oxford studies in metaphysics

volume 3

OXFORD STUDIES IN METAPHYSICS

OXFORD STUDIES IN METAPHYSICS

OXFORD STUDIES IN METAPHYSICS

Volume 3

Edited by

Dean W. Zimmerman

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Great Clarendon Street, Oxford OX2 6DP

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide in

Oxford New York

Auckland Cape Town Dar es Salaam Hong Kong Karachi
Kuala Lumpur Madrid Melbourne Mexico City Nairobi
New Delhi Shanghai Taipei Toronto

With offices in

Argentina Austria Brazil Chile Czech Republic France Greece
Guatemala Hungary Italy Japan Poland Portugal Singapore
South Korea Switzerland Thailand Turkey Ukraine Vietnam

Oxford is a registered trade mark of Oxford University Press
in the UK and in certain other countries

Published in the United States
by Oxford University Press Inc., New York

First published 2007

British Library Cataloguing in Publication Data
Data available

Library of Congress Cataloging in Publication Data
Data available

Typeset by SPI Publisher Services, Pondicherry, India
Printed in Great Britain
on acid-free paper by
Biddles Ltd., King's Lynn, Norfolk

ISBN 978–0–19–921839–4 (Hbk)
ISBN 978–0–19–921848–6 (Pbk)

1 3 5 7 9 10 8 6 4 2

PREFACE

Oxford Studies in Metaphysics is dedicated to the timely publication of new work in metaphysics, broadly construed. The subject is taken to include not only perennially central topics (e.g. modality, ontology, and mereology) but also metaphysical questions that emerge within other subfields (e.g. philosophy of mind, philosophy of science, and philosophy of religion). Each volume also contains that year's winner of the *Oxford Studies in Metaphysics* Younger Scholar Prize, described within.

D.W.Z

New Brunswick, NJ

This page intentionally left blank

CONTENTS

THE *OXFORD STUDIES IN METAPHYSICS* YOUNGER SCHOLAR PRIZE

Sponsored by the A. M. Monius Institute* and administered by the editorial board of *Oxford Studies in Metaphysics*, the essay competition is open to scholars who are within ten years of receiving a Ph.D. or students who are currently enrolled in a graduate program. (Independent scholars should enquire of the editor to determine eligibility.) The award is $2,500. Winning essays will appear in *Oxford Studies in Metaphysics*, so submissions must not be under review elsewhere.

Essays should generally be no longer than 10,000 words; longer essays may be considered, but authors must seek prior approval by providing the editor with an abstract and word count by 1 November 2007. To be eligible for next year's prize, submissions mailed within the United States or Canada must be postmarked by 15 January 2008. Authors mailing submissions from elsewhere should ensure that they arrive before 20 January. Refereeing will be blind; authors should omit remarks and references that might disclose their identities, but enclose a cover letter with contact information. Receipt of submissions will be acknowledged by e-mail. The winner is determined by a committee of members of the editorial board of *Oxford Studies in Metaphysics*, and will be announced in late February 2008. At the author's request, the board will simultaneously consider entries in the prize competition as submissions for *Oxford Studies in Metaphysics*, independently of the prize.

Previous winners of the Younger Scholar Prize are:

- Thomas Hofweber, ''Inexpressible Properties and Propositions'', Vol. 2;
- Matthew McGrath, ''Four-Dimensionalism and the Puzzles of Coincidence'', this volume;
- Cody Gilmore, ''Time Travel, Coinciding Objects, and Persistence'', this volume;
- Stephan Leuenberger, ''*Ceteris Absentibus* Physicalism'', forthcoming in Vol. 4.

* The A. M. Monius Institute is a non-profit organization dedicated to the revival of traditional metaphysics. Information about other activities of the A. M. Monius Institute may be found at <http://www.ammonius.org>.

Enquiries and submissions should be addressed to:
Dean Zimmerman, Editor
Oxford Studies in Metaphysics
Philosophy Department
Rutgers University
Davison Hall, Douglass Campus
New Brunswick, NJ 08903

Part I

SYMPOSIUM: HUMAN BEINGS

This page intentionally left blank

1. Human Beings, Human Animals, and Mentalistic Survival

Denis Robinson

INTRODUCTORY REMARKS

The debate over personal identity is a complex one, involving many contrasting views, and ingenious and controversial arguments. But standing back from the complexity, we can see a couple of major and strongly contrasting groupings into which many recently defended views may be seen as falling. At one extreme, we find diverse forms of Psychological Reductionism, and at the other, views we may group under the label ''Animalism''.[1] Though Psychological Reductionism is (deservedly, I think) more popular,[2] Animalism also has its following, and has fairly recently been given an unusually blunt, succinct, and passionate defense by Eric Olson in Olson (1997).

The doctrinal opposition between these main groupings tends to line up rather neatly with a general methodological opposition which has ramifications in metaphysics generally: one where the key issue is, roughly, the relative priority in metaphysics of folk intuitions and a priori judgments, as against the findings of natural sciences.

But the very distance between these two main doctrinal groupings makes it natural to look for some kind of intermediate position, and for some corresponding methodological middle way for defending it. My aim in this paper is to inspect this intermediate terrain, and in so doing to draw doctrinal and methodological conclusions of my own. My sympathies are with Psychological Reductionism, rather than with

[1] An account of the various positions mentioned in these opening paragraphs will follow shortly.

[2] Johnston calls ''dominant'' a genre of Psychological Reductionism which he names ''Wide Psychological Reductionism'', citing Sydney Shoemaker, Anthony Quinton, and David Lewis as amongst its defenders (Johnston 1987: 61). Olson (1997: 170 n. 13) lists over thirty philosophers, including nineteen whom he refers to (p. 20) as ''big names'', as holding some kind of Psychological view.

Animalism, but it is not my aim here to defend the former, nor to attack the latter. Rather, I aim to bring out the importance in this area of very general ontological assumptions or presuppositions, and the way in which the outcome of applying a particular methodology may be radically affected by them. I shall build my discussion around one well-known attempt—Johnston (1987)[3]—to defend a middle way between the extremes of Animalism and Psychological Reductionism. I shall argue that the attempt is unsuccessful, that the middle ground it attempts to occupy is unstable, and that the methodological proposals invoked in its defense cannot do what is asked of them.

Not only is Johnston's view in Johnston (1987) (henceforth, ''the Human Beings view''[4]) one which might be thought to combine elements characteristic of our main opposing camps, it also offers what might be seen as an attempt at a methodological middle way in support, since (as we shall see below) it invokes a priori intuitions about personhood on the one hand, while expressing caution about such intuitions, and prioritizing naturalistic ontological categories, on the other. I shall argue, however, that any substantive alternative to Psychological Reductionism which that methodology can be made to support will be even closer to Animalism than the Human Beings view, which is itself, despite the appearance of compromise, already quite close to an Animalist view.

To move on, we must begin by saying more about the kinds of positions referred to above. I shall say quite a bit about Johnston's position, the kind of ontological standpoint which underpins it, and the errors, as he sees them, which he wishes to avoid, before finally examining his methodological proposals, and the positive arguments for his view, to which they lead.

Psychological Reductionists agree that the constitutive criteria for personal identity over time are predominantly psychological. Various kinds of psychological connection may hold between a person's earlier

[3] It is important to realize that Johnston (1987) alone does not give an accurate impression of Johnston's overall position, as he in effect points out in Johnston (1987) n. 8, p. 64. Fully to appreciate his subtle and complex view one must read also Johnston (1989a), Johnston (1989b), and Johnston (1992b). Of these Johnston (1989b) is perhaps both the most important and the most difficult. A critique of Johnston's whole view would be a much larger undertaking than space here permits. My own take on the personal identity debate may be found in Robinson (2004).

[4] I capitalize the phrase ''Human Beings'' throughout, to signal that I am using it in Johnston's precisely articulated sense, not merely in an everyday sense.

and later mental states: combinations of such connections constitute the essential conditions for personal identity over time. Psychological Reductionists differ over the relevant connections and combinations, and their required causal underpinnings. (Johnston speaks of ''Wide Psychological Reductionism'', where ''Wide'' connotes a relatively liberal attitude to the permissible kinds of causal underpinnings.)

Animalism is the view that the familiar persons of our acquaintance are strictly and literally identical with members of a particular animal species, *Homo sapiens*. Members of this species are typically persons for most (but perhaps not all) of their lifespans. There is no special problem of *personal* identity: identity-conditions for a person of this familiar kind are just identity-conditions for the relevant kind of animal, much like the identity-conditions for other animals. An animal is born, it lives, it dies. For these events and processes, there are no essential mental prerequisites, certainly no requirement of psychological connectedness or continuity. Thus, Animalism is less a view about the nature of persons in general, more a view about the nature of those with which we are mostly acquainted. If there are angels, thinking automata, Martians, or dolphins which are people, then the identity-criteria for those persons will just be those appropriate to such beings: there are no identity-conditions for persons per se.

As noted, the doctrinal opposition between Animalism and Psychological Reductionism aligns naturally, though not inevitably, with a methodological opposition. Psychological Reductionist views are typically justified largely by appeal to intuitions about the loss or preservation of personal identity in various kinds of real and imaginary cases: by the ''method of cases'', as Johnston calls it. But this methodology has been called into question in a variety of ways.

Quine popularized the view that the analytic/synthetic distinction is at least hazy, and at best a matter of degree—some truths in which a term figures merely being more centrally constitutive of the term's meaning, or the concept it expresses, some less. This casts a shadow on any correlatively understood enterprise of conceptual analysis, of which the ''method of cases'' is typically a core component. The shadow falls particularly on our ability reliably to recognize a priori the genuine (''metaphysical'') possibility or impossibility of some circumstance—as opposed, for instance, to its degree of familiarity or bizarreness, or its relationship to well entrenched empirical assumptions. Complementing both this aspect of Quine's work, and his epistemological naturalism, we

have Kripke's and Putnam's work on rigid designation and natural kinds, suggesting that the essence of a kind of thing might need to be discovered by empirical inquiry, with conceptual analysis being, not merely epistemically suspect, but quite beside the point.[5]

These trends lead easily to the thought that we should reject the method of cases in favor of seeking where possible to identify the category of familiar persons with some prima facie appropriate natural kind. Biological kinds, and in particular animal species, are standardly taken as paradigms of such kinds. With a little help from evolutionary theory, Animalism emerges from these considerations as a natural rival to Psychological Reductionism.

In Johnston (1987) Johnston mounts a qualified critique of the method of cases, specifically in its application to the topic of personal identity, and proposes an alternative to Psychological Reductionism which, like Animalism, gives crucial precedence to the notion of a biological kind.[6] But he uses the notion of a biological kind in a more subtle way than does Animalism, in arriving at an account of the persistence-conditions for persons. Appealing to an alternative methodology, he argues that familiar persons are to be identified with ''Human Beings'': entities having persistence-conditions similar to, but not quite the same as, those of human animals. Indeed, though the kinds ''Human Being'' and ''human animal'' differ in their persistence-conditions, the crucial differences are rarely manifest in ordinary life, concerning as they do extreme and mostly imaginary circumstances.

Johnston's argument for this view appeals to the special role of certain psychological states, namely experiential memories, in our knowledge of personal identity; in this it is reminiscent of Psychological Reductionism, particularly in those neo-Lockean versions which make experiential memory crucial. Nevertheless, Johnston (1987) rejects the idea of straightforwardly mentalistic identity- or persistence-conditions. On the Human Beings view, psychological states and relations have a

[5] Though popular, the idea that Quine, Putnam, and Kripke provide us with reasons, good or bad, to replace conceptual analysis in the service of metaphysics with appeals to the findings and categories of the natural sciences is not inevitable. Frank Jackson's recent Jackson (1998) is one work which defends the view that these supposedly disparate methodologies should actually be seen as complementary.

[6] I shall not much discuss Johnston's attack on Psychological Reductionism as based on the method of cases. It revolves in part around a well-known conundrum for the method of cases due to Bernard Williams. For some criticisms of it, see Oderberg (1989).

merely evidential rather than a constitutive relation to personal identity. Any possibility that a person's survival might be sustained by psychological connections which transcend certain biologically based limitations, as in the fantasies of teletransportation, or of machine-assisted transfer of psychological states (or the informational states which realize them) from one brain to another, is ruled out. In all this, the view resembles Animalism. The key difference is that a Human Being, unlike a human animal, could persist through transplantation of its living brain into a different body, or in the form of the proverbial ''brain in a vat''.

''HUMAN BEINGS'', ''HUMAN ANIMALS'', AND THE CONSTITUTION RELATION

Johnston's view is that the (normal) relation between human beings and human animals is one of *constitution*, using this word in a technical sense commonly employed, with minor variations, by contemporary ontologists. The constitution relation holds between material things which occupy the same place at the same time and are composed of the same matter but which are nonetheless not strictly identical.[7] An important special case is the relation between a material substance and the matter which composes it. Notoriously, the constitution relation is a time-relative or changeable relation, since the matter composing material substances may change.

It's important to notice that the constitution relation may hold between material substances—bona fide, everyday *objects*—and not just between a substance and its composing matter. For instance, a mountain might be composed of a single gigantic rock; subsequently thermal stress may cause the rock to fracture into small pieces, yet the mountain persists. The mountain is at one time, not at others, constituted by the rock. Having different histories, they are never identical. Similarly, a cloak composed of a button and a piece of cloth may lose its button, which may be destroyed. Subsequently the cloak is constituted (solely) of the piece of cloth. Cloak and cloth are never identical, since at

[7] At least not in typical cases. There is a debate about whether identity can ever be a case of constitution. Cf. n. 9.

an earlier time the button is part of the cloak, though it is never part of the piece of cloth.[8]

Thus, the constitution relation is importantly different from identity, despite being readily confused with it.[9] An important instance of this general distinction, is that between cases in which one thing ''is'' another, in the sense of constituting it—as the boulder, at certain times, ''is'' the mountain, or the cloth ''is'' the cloak—from cases in which an entity falls, temporarily or contingently, under a ''phase-sortal'' predicate. The latter are cases of genuine identity. A person begins life as an infant, passes out of childhood into adolescence, and, with luck, grows out of adolescence into adulthood. For a limited time, the person ''is'' the adolescent. But this is an ''is'' of identity, rather than of constitution. The person *ceases to be* an adolescent, but is nonetheless *identical with* that adolescent. The adolescent, on ceasing to be an adolescent, does not on that account cease to be.[10]

[8] These examples are from Robinson (1982: 319). Though Olson has elsewhere contributed to the literature on the constitution relation, in Olson (1997: 101–2) he is remarkably curt and dismissive about views appealing to it. But if it is hard to deny there may be clear examples of ordinary things standing in this relation, appeal to it ought to be a prima facie option in the personal identity debate.

[9] Another point of which these examples may remind us is that, even after we set aside instances of the ''is'' of predication, it is neither necessary nor sufficient for the concept of identity to be invoked by a given statement that terms appear linked by ''is'' or other forms of the verb *to be*. In fact, I believe that even such everyday locutions as ''is identical with'' need not signal identity in the philosophers' sense: for they, too, may be used by the folk to express the constitution relation. It will be evident that I do not in general take speakers' introspective judgments as to their meanings or concepts as an infallible guide, since ordinary speakers will not automatically distinguish particular utterances which attribute the constitution relation from assertions of identity in the strict, philosophers' sense. Failure to note these points makes it spuriously easy to impute to common sense a strong commitment to Animalism.

As a four-dimensionalist I believe that the constitution relation can be analyzed in terms of identity of temporal parts: x constitutes y at t iff the appropriate temporal part of x is identical with the cotemporaneous temporal part of y. In special cases this means that identity is a limiting case of the constitution relation. Johnston however rejects such an approach (see Johnston 1992a). It follows that, on his view, no Human Being is ever identical with a human animal, nor with any temporal part of a human animal.

[10] For present purposes, we may characterize a ''phase-sortal predicate'' as follows. A ''substance predicate'', first of all, is one which applies to all and only members of a kind of entity, where members of a kind in the relevant sense share, inter alia, identity- and persistence-conditions. Applying a substance predicate to an entity thus carries implications for how it and members of its kind are to be individuated and tracked through time, and for which kinds of events they can and cannot, logically, survive. A phase-sortal predicate (such as ''adolescent'', ''blonde'', or ''widow''), finally, is logically equivalent to the conjunction of a substance predicate with attribution of some temporary or accidental characteristic, and thus, by the rules for conjunction, carries similar implications. See Wiggins (1980).

The important contrast here is with the case of the boulder which, I say, does indeed cease to exist when it fragments. Thus we should *not* see ourselves as referring to the mountain itself, under the description ''boulder'', when we refer to the boulder (though in many contexts it may be a matter of indifference which we refer to, so that we leave it indeterminate). Even more clearly, when we speak of the piece of cloth, in the example of the cloak, we are *not* merely speaking of the cloak itself under a phase-sortal description comparable with ''adolescent''.

Here I follow Johnston in adopting the vocabulary of what I call a *substance-ontology*. Wiggins (1980) is a well-known exposition and defense of principles typical of such ontologies (though with, inevitably, a number of idiosyncrasies and optional extras). According to these ontologies, the world contains a *limited variety* of *kinds* of *substances*. Each of these kinds has distinctive criteria associated with it, criteria which, if articulated, would specify what kinds of events could count as commencing or terminating the existence of members of the kind, and how members of the kind are in principle to be identified, distinguished, counted, and traced over time. Important amongst the substance kinds are the natural kinds, including the biological kinds. Substance-theorists who prioritize natural kinds typically view the associated persistence-conditions as naturally, as opposed to conventionally, determined.

Consistently with these views, substance-ontologists typically deny the reality, or disparage the status, of arbitrary mereological aggregates of substances or other entities. They also typically deny the reality of temporal parts of substances, and insist on sharply distinguishing issues of change and identity for substances from parallel issues involving either events, or portions or quantities of matter.

Substance ontologies may differ widely in the range of kinds they recognize, and how they see those kinds as related. Roughly speaking, we can distinguish the *form* of a substance-ontology, from its *content*—the particular view it takes as to what the constraints on substance-concepts, and on substances, are. Natural kinds will typically be included, but that leaves open what other kinds might also be admitted. Suitably well-defined or determinate conditions for identity and persistence will always be required, but there is much room for debate as to what ''suitably well-defined or determinate'' means here. But however these issues are resolved, questions about substance-identity will for substance-theorists always ultimately come down to questions about either nominally or

metaphysically essential properties of kind-members: those properties possessed essentially by things belonging to the kind in question.

Thus we may say that on the substance-ontology approach, persons will comprise all or some of the members of all or some of the kinds of substances which typically have what it takes to be a person. For present purposes, and compatibly with all Johnston says, we can sum up what it takes to be a person, as the capacity for a sufficiently rich, sufficiently integrated kind of mental life. Consistently with all this, one *might* take it that the kinds of substance there are include *none* whose persistence-conditions coincide with anything which could plausibly be considered as *psychological* unity-conditions—conditions for the unity and persistence of a mind considered as such. An upshot of such an attitude would be that ''person'' might not qualify as a substance-concept, nor even as a phase-sortal concept, being more akin to a mere general concept like ''fast mover'', or ''red thing''. Given such assumptions, Animalism is a natural (and ''naturalistic'') conclusion.[11]

Johnston adopts, not just the vocabulary, but the stance, of substance-ontology, as sketched above, including, it seems, the rejection of mentalistic persistence-conditions for bona fide substances.[12] Certainly his argument at a crucial point seems to rest on *some* restrictive set of background ontological assumptions—assumptions which effectively load the dice against any psychological account of personal identity. Johnston recommends a standpoint according to which psychological continuity and the like should be viewed merely as providing under normal circumstances relatively good evidence of personal identity,

[11] In addition to Olson, Animalists include Paul Snowdon, Peter van Inwagen, and David Wiggins. See Snowdon (1990), van Inwagen (1990), Wiggins (1980), and Wiggins (2001). Note that Wiggins's Animalism is different in its details from, for example, Olson's. Wiggins views the concept ''person'' itself as a substance-concept and indeed a natural kind concept, where the kind in question is precisely ''human animal''. Nevertheless, he has been prone to suggest, obscurely, that the psychological nature of persons might form part of the essence of this kind, and so have some bearing on the individuation of persons, thus diminishing slightly the distance between his form of Animalism, and a Psychological view of personal identity, but creating a degree of distance between identity-criteria for this kind, and for other kinds, of animal. But as compared to his view in Wiggins (1980), in Wiggins (2001) this distance from other Animalists seems to have been lessened.

[12] It should be emphasized that, as I understand it, this is no implication of substance-ontology per se, though it accords well with the pre-eminent status typically accorded by substance-theorists to natural kinds.

much as do fingerprints or physiognomy. It would be as serious an error to take normally reliable psychological criteria as *constituting* personal identity, as it would be to take sameness-of-fingerprints as constituting personal identity.

Suppose we ask: could a machine, by reading the information from the brain of a human person and depositing it in another brain (real or synthetic), thereby bring about a continuation of the person's mental life and so provide a means of continuing the original person in existence in another body? Johnston suggests an attitude to this idea, much like that which an Animalist might take. From such a standpoint, the suggestion would be as bizarre and loopy as the idea that a machine duplicating fingerprints, or a plastic surgeon reproducing physiognomy, could accomplish the same effect. The most charitable view we could take of such ideas about fingerprints or physiognomy would be that they depended on radically mistaken overgeneralizations. Reflection on what is actually *done by* a fingerprint-duplicating machine, or a plastic surgeon, should be enough to make the mistake plain, since *they disturb the very regularities which make this evidence normally reliable*. Similarly, Johnston suggests, with the information-transfer machine.

Johnston's views thus run along similar lines to those of the Animalist, however they diverge at a crucial point. His views are, (i) that the relationship between a human animal, and a Human Being (in his technical sense), is not the relationship of identity, but the relationship of constitution, and (ii) that a Human Being could *in principle* outlive a human animal which formerly constituted it, though this possibility is one which does not get actualized in real life as we know it.[13] Indeed we may pretty much say that according to the Johnston of Johnston (1987), every actual individual Human Being *is* a human animal—*provided* that we understand this ''is'' as the ''is'' of constitution.

PROJECTIVE ERRORS AND THE REJECTION OF THE METHOD OF CASES

Johnston does not think that the error of those who misguidedly understand as *constitutive* criteria what are merely normally reliable *evidential*

[13] Setting aside, I would presume, various gruesome and (literally) short-lived borderline cases involving beheadings.

criteria for personal persistence arises simply from egregious overgeneralization. The relevant overgeneralizations gain spurious plausibility, Johnston thinks, from our tendency to commit what he calls ''projective errors''. We are susceptible to a broadly Cartesian view of the self as a ''bare locus'' of consciousness: an enduring entity, distinct from anything physical, distinct too from any particular mental events, but underlying, sustaining, and unifying such events, simply by being a constant that-to-which-psychological-presentations-are-presented. No amount of physical continuity, nor even of psychological continuity, is absolutely necessary to the continued existence of such an entity. It is this view of selves as ''soul pellets''[14] which Johnston describes as a form of ''projective error'':[15] we are unaware of the physical underpinnings and complex processes which give rise to episodes of consciousness, and fall into the trap of projecting our lack of awareness into a lack in reality, taking it that there are no such underpinnings and complexities.

Johnston suggests that even those philosophers whose views about personal identity revolve around the explicit rejection of any such account of the self, nevertheless take seriously intuitions which amount only to silly overgeneralization, in part because they retain an unconscious residue of the soul pellets view.[16] For these and other reasons, Johnston proposes to reject the ''method of cases'', which endeavors to extract criteria for personal identity from, principally, our intuitions about real and imaginary examples of personhood and its vicissitudes.

REIDENTIFICATION, MENTALISTIC SURVIVAL, AND THE BRAIN AS THE NATURAL ORGAN OF MENTATION

Johnston does not totally renounce all appeal to intuitions about imaginary cases. But he advocates a strict methodological regime according to which only those intuitions which cannot be traced to misleading

[14] Cf. Johnston (1989a: 372). [15] Cf. Johnston (1992: 593).

[16] It's worth noting here that Locke's inauguration of the modern philosophical discussion of personal identity, used, in the specific context of an argument supposed to show the irrelevance of sameness of soul (or of organism) to personal identity, the example of the *soul* of a prince coming to inhabit the body of a cobbler. In other words, Locke might be accused of appealing to intuitions grounded in precisely the view he was attacking.

origins like those he identifies can lend support to an account of personal identity. Johnston's alternative proposal instead rests centrally on the observation that we normally identify and reidentify human persons relatively unproblematically, and proceeds by asking what kinds of entities *could* be unproblematically identified and reidentified in normal circumstances by the practices we habitually use.

Given this methodology, we might have expected Johnston to opt for Animalism, but in fact he opts for the Human Beings view. His argument includes two crucial steps. *First*, he asserts that "*if anything deserves the name of a conceptual truth about the relation between persons and minds, it is the claim that a person cannot be outlived by (what once was) his own mind*."[17] (I'll call this claim MS, for "mentalistic survival": it says *survival of the mind suffices for survival of the person*.) *Secondly*, Johnston urges "*a properly naturalistic view of our mental functioning*" as the characteristic functioning of a particular organ, the brain, claiming in this connection that "*talk of a mind is overly reified talk of an aspect of some minded thing*."[18]

These are the two crucial steps from which Johnston concludes that we should view ourselves as Human Beings, entities normally constituted by a human animal, but such that in extraordinary (and so far imaginary) cases such as brain transplants, or brains-in-vats scenarios, they might be constituted merely by a living human brain, or by a creature resulting from transplant of such a brain. On this view, the death not of a human *animal*, but of a human *brain*, constitutes the death of a human person.[19]

How, we may wonder, do these two steps mesh with Johnston's proposed alternative methodology? Even if MS and the conclusion he derives from it strike us intuitively as plausible, by Johnston's lights that in itself counts for little. Is his appeal to this principle really an improvement on a direct appeal to intuition? This is probably the point

[17] Johnston (1987: 77). Olson cites this remark (Olson 1997: 12) in a context which suggests he takes it to show that Johnston supports a Psychological account of personal identity—a debatable reading which seems to me to take the remark very much out of context. Johnston's account is "psychological" in what is at best a stretched sense, and he explicitly opposes typical psychological theories. Cf. n. 29, below.

[18] Johnston (1987: 78).

[19] In discussion, it is often suggested that Johnston's view here amounts to the view that a person is identical with their brain. This claim, which Johnston rebuts, certainly does not follow from the view that a person is *in certain extraordinary circumstances* to be traced by tracing their brain. The fact that my car could survive destruction of its doors does not mean that they are not really part of it!

where Johnston's argument seems most to resemble traditional appeals to conceptual analysis. Does he merely capitulate to the method of cases at this point?

MENTALISTIC SURVIVAL, EXPERIENTIAL MEMORY, AND NON-MENTALISTIC INDIVIDUATION

Johnston does attempt to tie his appeal to MS to his proposed alternative methodology. The link is meant to be the role of experiential memory in self-identification. To see how, we must examine MS and the use Johnston makes of it. To this end, let's first compare MS directly with comparable claims about body organs.[20]

The claim ''a person cannot be outlived by (what once was) his own *heart*'' is false: we are familiar with cases in which the heart remains alive and is transplanted into the body of another person, after the death of its previous owner. Does the same possibility exist, in principle, for the *brain*? Johnston urges, plausibly to my mind, that we could not say the human animal as such would survive if the brain alone were kept alive and transplanted.[21] So a human animal can be outlived by what once was its own brain. Now if the person were (identical with) the human animal, it follows that the person could likewise be outlived by what once was its own brain. The denial of this possibility, conversely, is the upshot of Johnston's argument in Johnston (1987). He does not deny the surgical possibility, but claims rather that survival of the transplanted brain *would* be sufficient for survival of the original person.

There is a peculiar subtlety involved in differentiating Johnston's view from Olson's particular version of Animalism, since Olson argues—contrary to the claim I made above—that an animal cannot survive without a brain, nor vice versa, taking the view that where the living animal brain goes, there the animal goes. So Olson does hold that an animal could in principle survive being reduced to the condition of a brain removed from the skull which once housed it, and kept alive

[20] A full discussion of whether MS is true or a priori would require a detailed discussion of what might or might not count as ''reidentifying a mind''. We may reasonably dispense here with such discussion because Johnston's next move substitutes, rightly or (as I think) wrongly, the issue of reidentifying a brain.

[21] But note that Olson disagrees on precisely this point. See Olson (1997: ch. 6).

through artificial means.[22] It may appear therefore that whereas Johnston takes persistence for Human Beings, on his account, to *differ* from persistence for human animals, according to Olson's account it precisely *coincides* with the latter!

This does not threaten the generic difference between the Human Beings view and Animalism, since Animalists might have other views than Olson's as to necessary and sufficient conditions for animal survival. Furthermore, though Johnston, as we shall see, associates human mental life with the human brain, Olson's view is that it is not the whole brain, but the cerebrum, which is ''that organ which is most directly responsible for your higher mental capacities such as reasoning and memory''.[23] If Olson is right about this, then Johnston should say ''*cerebrum*'' wherever he in fact says ''*brain*''. In addition, Olson does not consider it necessary that the human animal, to survive, should retain a functioning cerebrum—the remainder of the brain, if it is alive and performing its normal functions, will do. In other words, for Olson, the cerebrum is an *inessential* part of the brain qua organ essential to survival. Thus the appearance of coincidence between Olson's and Johnston's views disappears once we take these subtleties into account: the part of the human animal essential for survival of Johnston's Human Being is in fact disjoint from the part essential, according to Olson, for survival of the animal itself.

Finally, and most importantly, this coincidence would at best be only an extensional equivalence. Even were Johnston's reasoning and Olson's to arrive at exactly the same conclusion as to what it minimally takes for a human person to survive—as it might be, continued survival and functioning of the brain *as a whole*—their views would remain intensionally different, since they differ as to which functions of the brain would earn it this crucial status.

This is particularly relevant in the light of the fact that Johnston views the relation of human animal to Human Being as one of constitution, and in addition takes the view that constitution always excludes identity. It seems then that even if human animal and Human Being always exactly coincided, due to the brain's functional versatility, Johnston would be required by his own lights always to see them as distinct, whereas Olson would see the allegedly distinct substance-category of Human Beings as entirely spurious.

[22] See especially Olson (1997: 131–42). [23] Olson (1997: 9).

Returning to Johnston's main line of argument: he derives the claim that a person *can* survive through survival of their brain alone, from MS: the claim that it is a priori that survival of *mind* is sufficient for survival of person. And the second main step of his argument proceeds, essentially, by claiming that the truth of MS is preserved (though not, presumably, its a priori status!) if the word ''brain'' is substituted for ''mind''. He says in support of this: ''A human mind is just a mode of functioning of a natural unit (e.g. a human organism or a human brain) whose conditions of persistence are statable in nonmental terms. This is the sense in which talk of a mind is overly reified talk of an aspect of some minded thing.''[24]

Note how an implicit commitment to a restrictive ontology, most plausibly some kind of substance-ontology, emerges in the reference to *natural* units whose conditions of persistence are statable in *nonmental* terms. This is a pivotal point in Johnston's argument, and one at which his general ontological presuppositions come into view. Given the aim of discovering what kind of thing it is which we are most readily seen as unproblematically identifying and reidentifying in our everyday person-identificatory practices, it transpires that the candidate kinds are to be ''natural units'' such as organisms or body organs, or composite entities whose persistence-conditions may be formulated, *in principle non-mentalistically*, in terms of such natural units. Mentalistic persistence-conditions as such are ruled out.

Here a major general moral emerges. The results yielded by Johnston's methodology will inevitably be relative to the options the world is taken to serve up to us in the way of candidate kinds for our identificatory practices to be seen as tracking. Since different ontologies will take different stands on this, there is a danger that the choice of a particular ontology already tendentiously limits the outcome of the inquiry before any appeal to identificatory practices can enter the picture. I believe this actually happens in the case of the answer Johnston supplies. Indeed I believe that a central weakness of Johnston's preferred methodology is its inevitable need for some pre-established restriction and ranking of the kinds of entities which are to be seen as candidate targets of our identificatory practices. The more general moral, however, extends beyond the context of Johnston's own methodology.

[24] Johnston (1987: 80).

Consider for instance the defense of conceptual analysis mounted in Frank Jackson's recent Jackson (1998). Johnston's and Jackson's methodological proposals share a common structure. Faced with the metaphysical question "What are Fs?" each in effect proposes a test to be applied to the members of some favored basic ontological pool. Those members of the pool which best pass the test, thereby qualify to be counted as the Fs. Jackson's test may be thought of as generated by Ramsifying an implicit term-defining "theory" of Fs, taken as revealed piecemeal by conceptual analysis; Johnston's (in the specific case where the Fs are the *persons*), as generated by asking what kind of thing it is which we are most readily seen as unproblematically identifying and reidentifying in our everyday person-identificatory practices. Thus Jackson and Johnston fish with different nets. But what entities you can fish out of an ontological pool depends not merely on what net you use, but on what the pool contains in the first place. If mentalistically individuated entities are not to be found in it, no net will fish them out. In other words, any metaphysical methodology which has this broad structure, identifying the Fs by applying a net or filter to some predesignated ontological pool, will only be as good as the reasons available for preferring that pool to begin with. A central weakness of Johnston (1987) is the lack of any overt argument in favor of the preference, albeit a fashionable one, for a pool which contains biological and other natural kinds, but which excludes mentalistically individuated substances per se.

Johnston's choice here (to allow only non-mentalistically individuated entities as bona fide substances) is not altogether arbitrary or unmotivated. He claims that the method of cases fails in this context, in part because "the" concept of a person, taken as what is in common to all the different conceptions of personhood which have historically been held or which may emerge from reflection on imaginary cases and the like, is impossibly vacuous. Nothing but a soul pellet could fit it, yet a capacity unproblematically to reidentify soul pellets would be impossible to account for; nor would it make sense to attach the kind of importance to the vicissitudes of soul pellets which we in fact attach to issues of personal identity. Thus it is natural for him to assume that his alternative method must to some extent rely on *nature*, rather than on *our concepts*, to demarcate the persons: to assume, in other words, that the kind "Human Person" must be constructible relatively straightforwardly

out of natural kinds. And in this context, as noted, Johnston identifies the relevant natural kinds as the biological kinds.[25]

Nevertheless, it would be possible to meet the demand that the persons be *in some broad sense* "naturally" individuated, and yet provide a greater range of options in seeking a kind to match with our ordinary person-identificatory practices. Discounting hopelessly gerrymandered or arbitrarily or extrinsically individuated "kinds" need not oblige us to reject *mentalistic* or *psychological* persistence-conditions as Johnston does.

Johnston refers to "the naturalistic framework taken for granted by scientifically validated common sense": but I doubt that respect for scientifically validated common sense demands that we reject the possibility of mentalistic persistence-conditions. It is admittedly a matter of considerable philosophical debate how to integrate the mental into a naturalistic world view: but scientifically validated common sense is happy to admit all sorts of entities which are not, but which supervene on, those which figure explicitly in actual science. There is for instance no science of bricks *as such*, but given that bricks supervene on molecules, and are relatively naturally individuated, scientifically validated common sense has nothing to say against them. Similarly, I can commonsensically count the ripples on a pond and see how far they travel, with a clear scientific conscience, even if I acknowledge that ripples cannot exist independently of the quantities of water which successively, by their causally interrelated movements, constitute them. It's just not true that every bona fide materially constituted entity recognized by the union of common sense and science, and which persists and can be reidentified over time, counts as an independent substance which meets the stringent criteria laid down for substance-concepts by writers such as Wiggins, or which belongs to the biological natural kinds with which they are so taken.

In Johnston (1987) Johnston says nothing which specifically addresses this issue. Since Locke,[26] though, we should know better. Locke's most profound and original contribution, I believe, in his discussion of personal identity, is not his proposal of a psychological criterion for

[25] Johnston does acknowledge the need for further argument in support of his prioritizing of non-mentalistically individuated natural units, as preferred targets for our reidentificatory practices, in his n. 19 to Johnston (1987).

[26] All references are to Locke (1961, Book II, ch. 27).

personal persistence, but something which is logically prior to that proposal—his recognition of the fact that, given an ontology of fundamental natural kinds (even including ''supernatural'' kinds such as spirits), we may also recognize the existence of *additional* well-individuated substance-kinds, distinct from more fundamental and autonomous kinds, but at all times supervening on them, because constituted out of them. This is the lesson he draws from consideration of the case of living things, and applies, in a brilliant and original analogy, to the case of persons. It is a lesson of which materialists should stay specially cognizant, since the alternative historically was to believe in an entire category of fundamental immaterial substances, ''souls'' or ''spirits'', having their own primitive and unanalyzable persistence-conditions, and somehow guaranteed by their intrinsic nomic essence to sustain mental life and to individuate mental lives.

Locke in many ways fits the model of a substance-ontologist, attempting to argue for the cases of living things, and of persons, alike, that persistence-conditions for things of the relevant kind should be deducible from a careful statement of what it takes merely to qualify as a thing of that kind, and in effect arguing that in each case it is the persistence of an appropriate kind of *process* which is essentially constitutive of persistence for the relevant entities. Once again we see that the general form of a substance-ontology by no means automatically dictates the kind of prioritizing of biological kinds at the expense of mentalistic individuation, which Johnston must assume if his argument for the Human Beings view is to stand up.

Needless to say, there is no way fully to adjudicate the issues here touched on without a thorough discussion of the constitution relation, identity over time, and a host of other controversial topics in ontology.[27] Similarly, it would require a lengthy digression properly to defend my take on Locke. For my purposes it is sufficient to point out that Johnston is relying on assumptions which are by no means inevitable, which he does not defend, and which crucially determine the outcome of his argument.

[27] For instance, I am steering clear of the issue of four-dimensionalism which, as it happens, I endorse. Given unrestricted temporal partition and unrestricted mereological fusion, there will be no special problem about mentalistically individuated entities. But Johnston, as noted, in Johnston (1987) employs the language of substance-ontology. It is therefore dialectically more useful to point out that substance-ontology per se does not rule out mentalistic individuation, than to invoke the kind of apparatus—temporal parts and mereological fusions—which substance-theorists characteristically reject.

Let's return to that argument. It is not just outright Psychological Reductionism which Johnston renounces in renouncing mentalistic criteria for individuation or persistence, and the correspondingly richer set of kinds which might otherwise be regarded as eligible candidates for the title ''Human Persons''. Consider the intermediate or hybrid position I call the *Human Animal Stage* view. On this view, even if a person cannot be outlived by their own *mind,* their mind *can* be outlived by their body and brain. It holds that a person continues to exist so long as their mental functioning is substantively unimpaired, but that a person who suffers total permanent and radical amnesia and personality loss thereby ceases to exist. Subsequent development of new memories and personality will, on the Human Animal Stage view, constitute the coming into being of a new person, not the continued or renewed existence of the previous person.

Animalism differs from the Human Animal Stage view in denying that a person ceases to exist at such a moment of mental disintegration and amnesia. A Human Animal Stage theorist will hold that a single human animal may in such circumstances successively constitute two different people. But Johnston seems precluded from saying that radical psychological discontinuity would necessarily count as terminating a person's existence, or inaugurating the existence of another person, and hence (though in fact he does not consider such a view in Johnston (1987)) seems committed to rejecting the Human Animal Stage view.

A contrary impression may be given by the fact that Johnston tends to say merely that on his view a person *may* survive, *albeit in a mutilated form,* so long as their brain survives, and at one point in his introductory remarks he suggests a qualification, saying that ''a human being could be reduced to the condition of a mere brain *so long as* that brain continues the human being's mental life''.[28] In a similar vein, he says '' 'human being' names a partly psychological kind'', given that ''the tracing of a human being gives primary importance to mental functioning among the various life functions exhibited by human beings''.[29]

But Johnston repeatedly insists on giving the brain the basic individuative role when tracing Human Beings, as when he says ''it is crucial

[28] Johnston (1987: 64). My italics.

[29] Johnston (1987: 79). Johnston's words at this point are reminiscent of Wiggins's version of Animalism.

to the tracing of a human being that there be something that is the continued functioning of that human being's brain.''[30] In the light of this remark, it is clear that he cannot accept the Human Animal Stage view. The case raised above, in which the Human Animal Stage theorist would see a single human animal as successively constituting distinct people due to loss of mental continuity and connectedness, is one in which the brain does continue to function, hence by Johnston's lights must be seen as a case in which a single Human Being persists. Thus, Johnston should be seen as denying the necessity, as well as the sufficiency, of any sort of trans-temporal mental connections, for the survival of a Human Being. If mentalistically individuated entities as such are to be ruled out, we cannot make it any sort of condition for the survival of a Human Being *or* of its brain that the brain subserve any particular kind or pattern of mental functioning. The logical destination of Johnston's discussion is the unqualified claim that a person simply does survive just so long as their brain survives.

If this is the logical destination of Johnston's discussion there is something rather odd about his route to it. Let me explain.

Given his declared methodological assumptions, and his evident ontological assumptions, Animalism is and ought to be Johnston's natural starting point. The biological kind that at least our third-person, person-identificatory practices are most readily seen as tracking is the kind: Human Animals. But Johnston notes that we also habitually and unproblematically identify ourselves by the use of experiential memory. The deliverances of experiential memory, he says, are deliverances, true or false, about personal identity as such. As a fact of phenomenology, experiential memory always represents remembered experiences as experiences had by oneself: experiential memory does not deliver some ''halfway house'' judgment short of a belief about personal identity. Furthermore, its seeming aptness for delivering knowledge about personal identity would seem to survive even a Cartesian suspension of belief in one's body and in anything external to one's own mind.[31]

But if on the one hand we are already committed to finding a variant on the Human Animals view, in the sense of taking a *non*-mentalistically individuated biological kind of entity, or some composite constructed out of such entities, as the target of our ordinary modes of reidentification of

[30] Johnston (1987: 79). [31] See Johnston (1987: 77).

persons, whereas on the other hand experiential memory seems suited by its phenomenology to detecting merely *mental* connections, we seem in danger of having to conclude that it is after all merely minds, *not* persons, which experiential memory is best seen as unproblematically identifying and reidentifying. It *is* specially suited to detecting *mental* connections: it is *not* specially suited to detecting *physical* or *biological* connections.

Hence Johnston's appeal to MS. He needs this principle to rule out such an option. Any case of reidentifying a mind must, necessarily and a fortiori, be a case of reidentifying a person. And finally—to complete the resolution of this impasse—the putative substance-kind ''mind'' is revoked, and the organic kind ''brain'' is substituted for it.

In short, minds enter the picture because of the relative *un*suitability of experiential memory for the unproblematic identification of an organic entity such as a human *animal*. But once talk of a mind has done its job, the excessive reification is cancelled by reverting again to talk of an organic entity, but a different one—a human *brain*. The puzzle is that experiential memory has already been declared to be relatively unsuited—phenomenologically, at least—to reidentifying any such kind of thing unproblematically! I smell a rat.

In fact, it seems to me, Johnston's appeal to the role and peculiarities of experiential memory simply has the effect (unintended I am sure) of an exercise in misdirection. For all Johnston says, I believe, we might reasonably conclude that the things which can be ordinarily and unproblematically reidentified over time, in just the way in which we reliably and unproblematically reidentify ourselves and each other over time, are human animals. Johnston's preferred view gives a special status to the human brain, which provides the indispensable essence of a Human Being in his sense. I think it is important to see why his discussion of the role and character of experiential memory does not and cannot lend support to that conclusion. This is what I next attempt to show.

What is ''experiential memory'' best suited, phenomenologically, to reidentifying unproblematically?

Let's begin with the ''first-person'' phenomenology of experiential memory, and the claim that it represents an experience not simply as having happened, but as having been experienced by oneself. This

appeal to phenomenology seems to me at best a complete red herring. There simply is no logical room for an alternative phenomenology here: it follows just from the fact that *experience* is what experiential memory is memory of. One cannot recall an experience without recalling it as if it were one's own experience, because one cannot *have* an experience without having it as if it were one's own experience. A form of memory which informed one barely that some experience had occurred, without recreating that aspect of the phenomenology of experience, and *thereby* representing it as an experience had by oneself, would not be what we call "experiential memory". The phenomenology comes with the job description. What ought to determine the suitability of experiential memory as a normally reliable form of reidentification of a mind or of a self is something like the reliability of the information it delivers, interpreted as information about a mind or a self: not the form that information is packaged in. At least, not if the packaging is inevitable.

We can go beyond this criticism: further complexities arise concerning the phenomenology of experiential memory, obscure in their implications, relevant to Johnston's claims, but none of them supportive of his conclusions. At the very least, they highlight the lack of clarity of the precise claim which Johnston is making here.

For one thing, experiential memory very frequently provides us with information about the past states and behaviors of our *bodies*: the sleazy catch-phrase "let's go build some memories" suffices to remind us of this. Indeed, one might say that experiential memory frequently has a phenomenology which seems peculiarly well suited to *the reidentification of a body as such*. (Compare Descartes' Sixth Meditation remarks on the peculiar suitability of the faculty of imagination for providing us with information about the interactions of an *embodied* being with its environment.) Phenomenologically speaking, experiential memory often stops no more short of a judgment about the identity of a body than it does about the identity of a person (in contrast, be it noted, to the fact that it typically delivers nothing in particular in the way of judgments about the identity of *brains*).

For another thing, there is what I shall christen "the displaced viewpoint" effect. It may readily be confirmed by cross-examination that for many people, memory and imagination are alike in that remembering or imagining themselves performing some action often involves imagery of the action as seen from a displaced point of view. In remembering some particular occasion of going swimming, for instance, one may find

that what comes into one's mind is an image of oneself, seen from above, swimming in the water, rather than an image of the surface of the water as seen from an inch or two away, that is, from the point of view of the swimmer.

Now when one's memory takes such a form, it remains true that there is a particular first-person standpoint associated with the image: the remembered scene is viewed from a particular viewpoint. It could hardly be otherwise. The trouble with the displaced viewpoint effect, for Johnston's purposes, is not that in such cases experiential memory does not have an associated first-person standpoint, but that this standpoint is *not that of the person whose experience is being remembered*!

So it all depends on what we mean by ''phenomenology''. I may on one occasion remember seeing someone else swimming, actually having viewed them from above; or I might remember an occasion on which it was I who was swimming, but experience the displaced viewpoint effect, forming in my mind a memory image which duplicates that of the previous case. Typically, I will know perfectly well whether it is my own swimming, or my viewing of another's, which I am remembering. But I will not know this on the basis of the memory images per se. We have all learnt, after all, the Wittgensteinian lesson that imagistic accounts of thought are inadequate, since the content of such thoughts will depend crucially on the interpretation of the images, something which they themselves, qua images, cannot supply.

Perhaps, then, the proper way to understand ''phenomenology'' here is to take it as including, crucially, not just the character of memory images, but along with that, the *interpretation* of those memory images. Even so, given that memory images cannot supply their own interpretation, it's quite imaginable, whether or not it actually happens, that a person's experiential memory should be, in the following way, a little hazy: on a given occasion one might be confident enough one is remembering an actual event, yet, for all that, be in doubt whether the swimming figure in the memory image represents oneself, as in the displaced viewpoint effect, or someone else one actually watched swimming. Viewed in this light it's hard to see just how the phenomenology of experiential memory *does* fit it particularly well for the role of reidentification of selves or minds.

Perhaps this talk of memory images—or more generally, the idea that what is crucial here is recall of past experiences—is out of place. Despite his talk of phenomenology, Johnston himself says that the deliverances

of experiential memory are *propositions*[32] involving claims about personal identity. Perhaps his talk of ''phenomenology'' here is no more than dressed up talk of beliefs or judgments, and the claim is simply that usually when we remember some event, we remember that we were this participant rather than that? But this interpretation makes Johnston's talk of ''experiential memory'' and ''phenomenology'' inappropriate—I *might* find myself believing some proposition about my past actions in the absence of *any* particular phenomenology. Worse, it's hard to see how this claim on its own can carry the significance Johnston is seeking here.

At any rate, to repeat, it surely is not the *phenomenology* of experiential memory which makes it specially suitable for the reidentification of this, that, or the other, thing: it is the *reliability* with which its deliverances correlate with the facts concerning this, that, or the other, thing, which should be relevant here. Furthermore, as noted, Johnston himself says that what the phenomenology of experiential memory makes it specially well-suited for establishing is the existence of *psychological connections per se*—though he promptly transmogrifies this into the claim that it is only a mind per se which it is phenomenologically specially well-suited to reidentifying. And even this transmogrified claim cannot survive his subsequent, naturalistically motivated substitution of ''brain'' for ''mind''. (This is arguable, it seems to me, even if we set aside any merely generic objection to apparent substitutions within intentional contexts, which is what claims to know identity, whether of mind or of brain, surely are.)

The contingency of experiential memory as a means of self or *brain reidentification*

In the light of these considerations, and of Johnston's own insistence that we should not too hastily promote normally reliable evidential criteria into metaphysically essential constitutive principles, let's reflect further on the *contingency* of the fact that this form of memory normally gives us reliable information as to who we are, or indeed of any of a number of related facts. If someone is having a putative experiential memory, there are a number of questions which may be raised about it.

[32] Johnston (1987: 77). My italics.

a) Did an experience such as that which is apparently being remembered ever *actually occur*, and *cause* the apparent memory?
(A putative experiential memory cannot be genuine unless it sufficiently matches some experience, in the sense of providing sufficient true information about the character of some real experience.)
b) Supposing (a) answered in the affirmative, did the experience which sufficiently matches the memory, and which had a role in causing it, have the *right kind of role* in causing it, for the putative experiential memory to count as genuine?[33]
c) Did the same *person* have the original experience, as is now having the experiential memory?
d) Did the same *human animal* have the original experience, as is now having the experiential memory?
e) Was the original experience realized by states of the same *brain* as the brain whose states are now realizing the having of the experiential memory?

How these questions are related to one another will be a matter for dispute between those who hold differing accounts of personal identity. But I claim that it ought to be agreed *by all parties* that in any given case of someone having a putative experiential memory, it is only contingent on certain in-principle-alterable facts about the world as we currently know it that *any* of them may be answered affirmatively. This ought already to be obvious from the fact that people are capable of psychotic delusions in which they imagine themselves to be, for instance, Napoleon. It follows again from the possibility in principle, exploited in such works of fiction as the movie *Total Recall*, starring Arnold Schwarzenegger, of technology capable of implanting false memories. Experiential memory indeed gives those of us who are not psychotic a highly reliable source of knowledge of who we are. But this is contingent on facts which could in principle change.

[33] If A has an experience and describes it, and B hears, reads, or encounters that description and is thereby caused to have an apparent memory matching A's original experience, this is not a case of genuine experiential memory, whether or not B is, as in the case of reading one's own old diary entries, identical with A. Which other causal connections would or would not be sufficient for genuine experiential memory must remain to some extent moot here, but clearly we must count the normal neurophysiological processes which underlie the formation of experiential memories in the human brain as sufficient.

Suppose we set aside the issue of psychotic delusions and other apparent memories which are false or fictitious in the sense implied by a negative answer to question (a). We might imagine, for instance, that we come to know that the world has been made free, through hand in hand advances in technology and social norms, from psychotic delusions, the depradations of stage hypnotists, spurious cases of ''recovered memory syndrome'', and the like. Consistently with that assumption, a variety of further scenarios are possible.

First case. Suppose that it becomes technologically possible to perform brain transplants without damage to the normal processes by which memory-traces are laid down and preserved; and suppose that in other relevant respects not so far touched on, the world goes on much as it does at present. Then someone's having a particular experiential memory *would* normally provide reliable evidence that the brain whose states realized the original experiences was the same *brain* as the one whose states were currently realizing those experiential memories. At the same time, it would (*pace* Olson's views about animal persistence) provide somewhat less reliable evidence for identity of the relevant human organism if any. This is roughly in accordance with Johnston's view, though of course the memories would not have the evidential status just described, for the particular person *having* those memories, *unless that* person was in possession of the relevant facts—unless they actually *knew about* the existence and functioning of brains, the possibility of transplanting them, and so forth. But we can imagine other scenarios.

Second case. Imagine instead that the technology of brain transplants stays in its current undeveloped state, but that the technology of prosthetic brain implants takes off. These implants, we might imagine, substitute modules relying on silicon chip technology for groups of brain cells, relevant information being copied from the original brain cells, stored temporarily, and then entered into the chip cluster with which those cells are replaced. Eventually an entire brain can be replaced in this manner. Thus degenerative brain diseases like Alzheimer's may be defeated. Under such circumstances, experiential memory would provide a *more* reliable indicator of identity of human organism than of identity of brain: though once again, only those in the know about the relevant facts would be able to *employ* it as such an indicator. (Olson, once again, would dissent: his view is that under these circumstances the human organism would no longer exist.)

Third case. Imagine, finally, that technology was created capable of recording information from one person's nervous system and implanting it in the form of experiential memories in another's. It would then be entirely in order to accept the veracity of an experiential memory in the sense of believing it to be faithful to some actual experience—always to answer question (a) in the affirmative—whilst at the same time doubting or wondering whether it was indeed one of *one's own* experiences. If use of such technology was rife, the easy and uncomplicated ways in which we use experiential memory in tracing people would no longer be adapted to tracing things of that kind, and this would equally be the case whether one held the Human Animals view, Johnston's own Human Beings view, or even—it is important to note—some reasonably sophisticated version of Psychological Reductionism. An affirmative answer to question (a), even an affirmative answer based on experiential memory, would provide no reliable basis for answering questions (c), (d), and (e)—questions about identity, respectively, of experiencing person, of experiencing animal, and of experiencing brain.

In fact, of course, none of these possibilities obtains. As things stand, the actual role of experiential memory in unproblematically reidentifying persons accords well enough with both the Human Animals and Human Beings views of what a person is, but on each of these views its reliability in this role must be seen as contingent and a posteriori rather than as necessary or a priori. There are scenarios whose impossibility is merely technical—that is to say, scenarios which are not even nomically impossible, let alone metaphysically impossible—in which the special suitability of experiential memory for reidentifying either, or both, of these kinds of entities would be undermined. Furthermore, as things stand, experiential memory is also equally good for reidentifying persons conceived as the Psychological Reductionist conceives them!

CONCLUDING REMARKS

The upshot is that, as I see it, there is no way to combine the claims

(i) that *experiential memory is specially well-fitted to identifying minds*;
(ii) that *as a matter of a priori necessity, minds can't outlive the persons whose minds they are*; and

(iii) that *the natural unit whose functioning underlies the phenomena overly reified as "minds" is the brain*, to arrive at the conclusion

(iv) that *experiential memory is specially well-fitted to identifying persons, conceived as entities whose persistence-conditions are derivative on those of brains, as opposed to persons conceived as human animals*.

Claims (i) and (ii), if true at all,[34] are true only of minds qua minds: substitute "brains" for "minds" and any special suitability of experiential memory for the job of reidentification is undermined, and with it, any specially compelling reason based on that suitability for abandoning the Human Animals view in favor of the Human Beings view.

Experiential memory is, as (i) asserts, indeed well-suited *as things actually are* for the unproblematic reidentification of minds. Even this, I believe, is quite contingent. Still, if any of these kinds of entities—minds, human brains, human animals, Human Beings—may be seen as distinguished by the special suitability of experiential memory for unproblematically reidentifying its members, it will be the category of minds *as such*. And if as good materialists or naturalists, we really want to avoid reifying minds, we should here introduce as our surrogate not the materialistic category of brains (nor cerebra!), but the category of *mentalistically individuated persons*, conceived as Psychological Reductionists conceive them: possibly discontinuous physical entities which realize various mental states at various times, but such that those mental states collectively exhibit both synchronic and diachronic unity of a kind and degree appropriate to personhood.

Now of course such entities cannot be acknowledged if mentalistically individuated entities are ruled ineligible at the outset. Nor can they be acknowledged if one assumes a background ontology which is in some other way too restrictive. In principle such entities could coincide with mere temporal parts of human animals; numbers of them could be alternately and discontinuously constituted—or continuously and multiply constituted—by human animals, as in Jekyll and Hyde and other multiple personality cases; they could undergo fission and fusion; they could be discontinuously constituted first by some human animal, later by some android; and so forth. I've never found a convincing

[34] In what sense (ii)—which is to say, MS—is true, if any, is an interesting question, but one which space does not permit me to examine further here.

argument purporting to establish that such entities cannot be acknowledged consistently with the main tenets of substance ontology, provided that mentalistic individuation, which Johnston baulks at, is seen as acceptable. Nevertheless, substance ontologists (in line with their tendency to extend exclusive special privileges to natural kinds) are typically reluctant to admit such entities into their ontologies (though four-dimensionalists would have no problem with them).

It has by no means been my purpose here to *defend* Animalism, even though I believe that Johnston's argument in favor of his Human Beings modification of it does not succeed. Rather, my aim has been to point out that Johnston's sole argument in favor of modifying the Human Animals view does not stand up, and in particular that Johnston's proposed methodological alternative to the method of cases provides no good grounds for preferring his ''Human Beings'' modification. First, his methodology does not actually rule out his principal target, Psychological Reductionism, without having to be bolstered by an unargued and, I believe, indefensible ban on mentalistic individuation, or some equivalent form of ontological conservatism. Without good arguments for such assumptions, which he does not supply, Johnston's argument is essentially question-begging. But secondly, *even if* some suitably conservative ontological perspective is adopted, Johnston's methodology does not actually favor the Human Beings view over the Human Animals view. The kind ''human animals'' is certainly a more natural kind than the kind ''Human Beings'' which, as he admits elsewhere,[35] could include a creature consisting of the living body of a tiger onto which a living human brain had been successfully transplanted.

It is noteworthy, finally, that Johnston should purport to derive from his natural-kinds-and-normally-reliable-evidence methodology a conclusion which it does not favor, though it is relatively strongly favored by the alternative methodology of appeal to normal intuitions, which Johnston rejects. Derek Parfit, for instance, says ''suppose that my brain is transplanted into someone else's (brainless) body, and that the resulting person has my character and apparent memories of my life. Most of us would agree, after thought, that the resulting person is me.''[36] I think it is easy to go along with Johnston's argument because of these intuitions, which I am perfectly happy to accept. The point is that Johnston himself, officially, has no right to them.

[35] Johnston (1989b: 456). [36] Parfit (1971: 4).

In a nutshell, Johnston's alternative methodology asks the following questions. What sorts of characteristics do we take as providing good ''fingerprints'' for persons? And what kind of thing is it which such characteristics would indeed provide good fingerprints for? Literal fingerprints work well, as we all know. Experiential memory does too. But experiential memory provides a reliable fingerprint for a human brain—hence for any kind of entity of which a human brain is an essential part—only contingently. And under the contingencies which in fact obtain, it provides exactly as good a fingerprint for a human organism. In one respect, indeed, it provides a better one. One needs much less special knowledge—in particular, one needs much less knowledge of human anatomy—to be in a position to know of the general reliability of experiential memory as a fingerprint for human organisms, or to be in a position to know that one is dealing with an exceptional or unreliable case—than one needs in order to know of the general reliability of experiential memory as a fingerprint for human brains, or for entities whose nominal essence includes them.

The normal intuitions about brain transplants reflect, I believe, not a view about which kinds of fingerprints provide the most reliable means of identifying which kinds of entities, nor merely a residue of bygone philosophies or religions, but our sense that what we really *care* about in caring about personal survival is psychological unity, so that we prefer *where possible* to identify personal unity with psychological unity. I believe, but shall not argue here, that this is a preference which neither the world, nor our conceptual commitments, once shorn of false metaphysics, prevent us from indulging.

REFERENCES

Jackson, Frank (1998) *From Metaphysics to Ethics: a Defence of Conceptual Analysis* (Oxford: Clarendon Press).

Johnston, Mark (1987) ''Human Beings'', *The Journal of Philosophy* 84: 59–83.

——(1989a) ''Fission and the Facts'', *Philosophical Perspectives* 3: 369–97.

——(1989b) ''Relativism and the Self'', in *Relativism: Interpretation and Confrontation*, edited with an Introduction by Michael Krausz (Notre Dame, Indiana: University of Notre Dame Press).

——(1992a) ''Constitution Is Not Identity'', *Mind* 101: 89–105.

——(1992b) ''Reasons and Reductionism'', *The Philosophical Review* 101: 589–618.

Locke, John (1961) *An Essay Concerning Human Understanding*, ed. J. Yolton (London: Dent).

Noonan, Harold (1993) ''Constitution Is Identity'', *Mind* 102: 133–5.

Oderberg, David (1989) ''Johnston on Human Beings'', *The Journal of Philosophy* 86: 137–41.

Olson, Eric T. (1997) *The Human Animal: Personal Identity Without Psychology* (Oxford: Oxford University Press).

Parfit, Derek (1971) ''Personal Identity'', *The Philosophical Review* 80: 3–27.

Robinson, Denis (1982) ''Re-Identifying Matter'', *The Philosophical Review* 91: 317–41.

Robinson, Denis (1985) ''Can Amoebae Divide Without Multiplying?'', *Australasian Journal of Philosophy* 63: 299–319.

——(2004) ''Failing to Agree or Failing to Disagree?: Personal Identity Quasi-Relativism'', *Monist* 87: 512–36.

Snowdon, Paul (1990) ''Persons, Animals, and Ourselves'', in C. Gill (ed.) *The Human Person and the Human Mind* (Oxford: Clarendon).

van Inwagen, Peter (1990) *Material Beings* (Ithaca: Cornell University Press).

Wiggins, David (1980) *Sameness and Substance* (Oxford: Blackwell).

——(2001) *Sameness and Substance Renewed* (Cambridge: Cambridge University Press).

2. ''Human Beings'' Revisited: My Body is Not an Animal

Mark Johnston

Twenty years ago, in the paper ''Human Beings'' and elsewhere, I defended an alternative methodology for arriving at an answer to the question: What kind captures our essence and so determines our conditions of survival over time?[1]

Previously, when it came to philosophical theorizing about personal identity, the popular methodology—''the method of cases''—had been to collect ''intuitions'' about real and imaginary cases of personal survival and ceasing to be, and then bring those intuitions into some sort of reflective equilibrium that bore on the question of the necessary and sufficient conditions for an arbitrary person's survival. Imagined cases were treated as more or less on a par with real cases; for the then natural idea was that we should not restrict our evidence base to the adventitious experiments of step-motherly nature, when we could also avail ourselves of the ingenious thought experiments in the philosophy journals.

1

There are a number of reasons why this simple method of cases, with its implied parity of the real and the imaginary, should be rejected. First, the specific necessary conditions on our survival, conditions that are the upshot of our common essence, need not be available to armchair, or ''a priori'', reflection. Certainly, it would be strained to require that we must grasp such de facto necessary conditions in order to count as

[1] ''Human Beings'', *Journal of Philosophy* 84 (1987), reprinted in M. Tooley (ed.), *Metaphysics* (Garland Press, 1993); J. Kim and E. Sosa (eds.), *Metaphysics: An Anthology* (Basil Blackwell, 1999); and *Postgraduate Foundation in Philosophy* (Open University Press, 2002).

semantically competent with terms or phrases like ''person'' or ''is the same person''. That thought generalizes to put pressure on a once popular model of philosophical analysis as the teasing out and codifying our semantic competence. This model, which makes philosophy look and feel like a sort of advanced lexicography, might be replaced with the idea of philosophical analysis as aiming for the real definition of the item or phenomenon in question. This involves using all of the relevant knowledge and argumentative ingenuity we can muster in order to say *what it is to be* the given item or phenomenon. Clearly, much of that knowledge will be a posteriori, even if not particularly recherché.

A large part of my reason for returning to the topic of personal identity after so many years is to provide an illustration of the viability of the method of real definition.

Real definition is not inherently at odds with conceptual analysis. Concepts are themselves among the legitimate targets of real definition. For given any complex item, we can enquire after its constituents, and the relevant manner of combination that yields the complex in question. So a concept may be a conjunction of other concepts, or a quantification over other concepts, or built up from other concepts in still more detailed ways. Articulating any such conceptual structure looks as though it deserves the name ''conceptual analysis'', even though we now know that the results of cognitive science and empirical linguistic semantics will properly condition most of the interesting claims about conceptual structure. Even in the case of the analysis of (read: the real definition of) concepts, a posteriori matters of fact will play a crucial role.

Still, it will be urged that the application conditions of a concept must be at least implicitly understood by possessors of that concept, unless those possessors are disposed to defer either to experts (as with the concept of a tort in law) or to reality itself (as with the concept of water) to settle the extension of the concept. Just so; none of that need be denied in order to put real definition at the center of the philosophical stage. When the pattern of deference goes all the way to reality itself, as with the concept of water, then, in determining the extension of the concept across possible situations, we would do well to note our partial understanding of the concept, and get on with the inquiry into the real definition of the phenomenon it picks out. So when investigating the concept of water and its extension across possible situations, we quickly find ourselves considering the question of what it is to be the stuff water,

and whether, for example, the manifest form of water enters into its nature or essence along with its chemical composition, or is instead no more than a contingent appearance of water.[2] So if, in the same way, we defer to reality in our use of the concept of personal identity then we would do best to go straight to the real definition of a (human) person.

Even so, it may plausibly be held that in possessing the concepts of a person and of personal identity we are not deferring to experts, or even to a reality that is beyond the ken of ordinary users of these concepts. We all know well enough what persons are! This may be the best remaining defense of the old a priori method of aiming to articulate the application conditions of the concept of personal identity simply by considering our intuitive judgments in imagined cases.

Against this defense of the method of cases, ''Human Beings'' argued that our common *concept* of a person is ''highly determinable'', a polite way of saying that it does not carry very much content. This is shown first by the enormous range of detailed and conflicting conceptions (theories, ideologies, theologies) of persons and personal survival that have been expressed by using the common concept. Secondly, when we let the concept of a person run free in the full range of imaginary cases, the resultant picture of what it picks out is that of a ''Bare Locus'' of mental life, something that in principle could survive any amount of psychological and physical discontinuity. It is quite implausible to suppose that we are tracing any such things when, in the easy and offhand ways that we do, we trace ourselves and others through time.

So the method of cases, understood as a way of articulating the application conditions of our concept of personal identity is, at the very least, in need of reining in by a more robust sense of what is really possible in the way of human survival. Where could that come from if not from the real definition of a (human) person, an account of what it is to be a (human) person? Hence the somewhat non-standard way of putting the question of personal identity: What kind captures our essence and so determines our conditions of survival over time?

2

There is a second worry about the appeal to the method of cases in articulating the necessary and sufficient conditions of the survival of

2 For more on this approach see ''Manifest Kinds'', *Journal of Philosophy* 94 (1997).

something over time. We do not find much evidence that in tracing objects and persons (as opposed to things in other categories, such as performances, for example) through time we are actually deploying knowledge of *sufficient* conditions for cross-time identity. Instead, as a matter of empirical fact, it appears that nature saves us cognitive labor by having us ''offload'' the question of sufficiency onto the objects and people themselves—if I may put it that way. Objects of various kinds are salient to us, they attract our attention, and we trace them through space and time. As long as they do not manifest changes in respects we know to be important for their kind, we are ready to credit them as having survived, even if we remain properly agnostic about what their persistence actually consists in. So we should be prepared to discover that in tracing objects we are deploying knowledge of *some* necessary conditions for their survival over time—the thing can't explode into smithereens, for example—but not of any non-trivial sufficient condition. The objects just take care of themselves in this regard, they either persist or cease to be; to witness such outcomes we need not know any sufficient condition for their persistence.

If this plausible empirical conjecture as to ''offloading'' turns out to be true, it should be very bad news for those philosophers who suppose that in tracing objects or persons we are in the cognitive situation of having to stitch together short-lived entities like phases or temporal stages into significantly persisting wholes by means of ''gen-identity'' relations. A gen-identity relation is a necessary and sufficient condition for two temporal stages (or phases, or temporal parts) to be stages of the same persisting object. But as to the sufficient conditions, they are precisely *what we do not need to know* when we are tracing salient objects which are naturally individuated anyway.

An example might bring out the point more clearly. Here I confess immediately to using an imaginary case, but only in the service of clarification. (I also hope the example shows that rejecting the parity between real and imagined cases doesn't have to be a puritanically *joyless* exercise; we need not banish the small pleasures of cartoon-like science fiction from the topic of personal identity. Everything depends on how the cases are used.)

Suppose we are in a civilization where androids are, on the surface, and in terms of speech and behavior, indistinguishable from humans. Unlike ours, however, their heads just contain air-conditioners; that is, their heads contain ''organs'' for cooling their bodies. (Aristotle's view

of us would actually be true of them.) The androids' heads, unlike ours, are not at all where the action is when it comes to their survival over time. An android can function very well for days without a head. Everything important is going on in the torso of the android. Now suppose that you have become familiar with someone, whom you take to be an android, while he is in fact a human. And suppose you have been unwise enough not to rely on offloading, that is, on just letting this salient person capture your attention as he *does in fact* persist through time. Instead, you concentrate on short-lived stages of your acquaintance, and knit them together in terms of some criterion of identity; which, because you take him to be an android, amounts to something like: one stage of such a being is a stage of the very same being *if* and only if there is a pattern of continuity connecting the torsos of the stages.

Then, the inevitable takes place. Your acquaintance loses his head. But the torso, or "head complement", remains intact. Even though we have stipulated that your acquaintance is human and has not survived, you will count him as surviving, at least until the manifest signs of death are undeniable. For the gen-identity criterion you were using knits together "headless torso" stages with the earlier stages of your acquaintance. For you have been treating continuous preservation of the torso as sufficient for identity.

In doing this you are led to the wrong judgment; in fact, your acquaintance does not survive the destruction of his head. He is a human, not an android.

You would have been much better simply waiting and seeing just what kind of thing he was. You had no business deploying a sufficient condition for his identity over time. In fact, your using that sufficient condition did not help you trace him forward in time. All the previous evidence for his survival, the evidence of gross bodily integrity and consistency of dispositions to speech and behavior, would have been just as good evidence that he survived. You did not need to settle what exactly he was in order to trace him through time.

Your fault was unnecessarily to settle in advance, a priori as it were, what kind of thing he was. You could have "offloaded" that cognitive task onto him; his nature, which you do not fully know, settles what kind of thing he is.

My hypothesis is that our cognitive system does "offload" in precisely this way. For the very simple purposes of tracing and re-cognition, we

need not know, and do not know, sufficient conditions for identity over time. The developmental facts, concerning among other things the early age at which we in fact trace objects and persons, seem to support this hypothesis.

Notice that something about the case of the ''android'' who was really a human suggests that we can be aware of persisting items and trace them through time even if we are ignorant of the sufficient conditions for their identity. In the case of the ''android'' who was really a human, you made *a mistake about your acquaintance*. That implies that you still had your acquaintance as a topic of your thought and talk, and that you were able to trace him through time, even though you were ignorant of the sufficient conditions of his persistence through time. This suggests that tracing objects through time does not involve drawing on sufficient conditions for their persistence.

Trying to make our implicit knowledge of sufficient conditions for identity explicit by examining our reactions to merely possible cases will be a particularly pointless task if the hypothesis of offloading is correct.

3

Both ''Human Beings'' and ''Reasons and Reductionism'' played up a third worry about the implied parity of real and imaginary cases.[3] In the massive core of real cases, many sources of evidence for personal survival, such as persistent bodily integrity and mental continuity, converge and agree; whereas the whole philosophical charm and utility of the imagined cases in the literature on personal identity lies precisely in teasing these elements apart. The obvious question arises: Might we not have thereby *undermined* our ability to make good judgments about personal identity when considering these very cases?

That question acquires a sharper bite if we make the distinction not exactly between real and imaginary cases, but between the massive core of *ordinary* cases in which bodily, behavioral, and mental evidence converges, and the *fringe* cases, real or imaginary, where those sorts of evidence come apart. Bear in mind the moral horrors that await many of us around the end of our lives, thanks to the perverse incentives of modern medicine; mental death, coma, and absurdly protracted

[3] ''Reasons and Reductionism'', *The Philosophical Review* 101 (1992).

languescent shade-life. Who is to say that such all-too *real* conditions present very much simpler objects of philosophical diagnosis than Derek Parfit's most outré *imagined* cases? The same goes for pre-fetal matters. Neither the beginning nor the end of life present the best starting points for a theory of personal identity. It's in the massive core of ordinary cases of viable healthy human life where we can best trust our judgments of personal identity.

That the relevant distinction is between ordinary and fringe cases and not between real and imagined cases should have been obvious upon reflection. For imaginary cases can sometimes presage real cases; our imagination is sometimes the source of future reality. But where the imagined case is also a fringe case, it will remain so even if it becomes a real case.

In order to press home the worry about the alleged parity between ordinary and fringe cases, ''Human Beings'' noted three potentially distorting influences that could condition our judgments in fringe cases. The distorting influences were the psychological and social continuer effects, and the effect of a quasi-Cartesian illusion that we are only essentially just that of which we can be certain of being.

It is possible to imagine cases in which a person ceases to be, but nonetheless has an excellent continuer of his psychological life. In my view, teletransportation, in which a person's body is destroyed and immediately copied at a remote location, is just such a case. The person who appears at the receiving end remembers the acts and the experiences of the previous person, he will carry out that person's intentions, and he has beliefs and desires that almost precisely correspond to those of the person he replaced. Yet, those of us who think of teletransportation as a method of copying people which has the singular defect of destroying the original copy, will have to admit that teletransportation would serve up, at least on the mental side, the best sort of evidence we usually have for personal identity. There is therefore, in such imagined cases, a natural tendency simply to accede to such evidence, and trace a person through the events which preserve such psychological continuity. This I dubbed ''the psychological continuer effect''. It could very well just be an understandable overgeneralization from the ordinary run of cases, where such psychological evidence is highly correlated with the preservation of the bodily integrity of the person in question.

A further possible distorting factor in the fringe cases—''the social continuer effect''—arises from our practical interests in the relation of

personal identity. In everyday life the persistence of a person typically guarantees that over the short term there will continue to be some occupant of the particular complex of detailed and manifold social roles (lover, friend, leader, supporter, colleague, nemesis, regular customer, etc.) which made up that person's social life. Typically, over short periods, the holding of personal identity guarantees that one will have a unique social continuer, namely oneself. And the holding of personal identity is typically the only explanation of social continuation. So, reinforcing the psychological continuer effect we have the tendency to trace a person in terms of his social continuer. This will be a distorting effect so long as we are not essentially occupants of complexes of social roles.

When I wrote ''Human Beings'' the dominant position in the philosophy of personal identity was the so-called Wide Psychological View; which has it that a person survives some event or series of events just in case he has a unique psychological continuer around after the event.[4] In order to defend the Wide Psychological View, a theorist *needs* to appeal to intuitions about fringe cases, cases like teletransportation that are designed to tease apart bodily and psychological continuity. Otherwise, he has no argumentative leverage against views which require some form of bodily continuity in order for personal identity to hold.[5] Any such defense of the Wide Psychological View must face the question of whether the crucial intuitions are simply driven by the psychological and social continuer effects. Which makes the Wide Psychological View quite hard to defend.

Perhaps the best thing we can say for the Wide Psychological View is that it gives roughly correct conditions for the persistence of

[4] See Lewis, ''Survival and Identity'', in Amelie O. Rorty, (ed.), *The Identities of Persons* (Berkeley: California UP, 1976); Quinton, ''The Soul'', *The Journal of Philosophy* 59/15 (19 July 1962), 393–409; S. Shoemaker, with Richard Swinburne, *Personal Identity* (Oxford: Blackwell, 1984). Shoemaker clearly had some sympathy for the view that bodily continuity is constitutive of personal identity when he wrote *Self-knowledge and Self-identity* (Ithaca, NY: Cornell, 1963), although he allowed that in exceptional cases the bodily criterion of personal identity could be overridden by the memory criterion. He decisively abandons the bodily criterion in ''Persons and Their Pasts'', *American Philosophical Quarterly* 7/4 (October 1970), 269–85.

[5] There is one exception to this claim. In ''Personal Identity: A Materialist's Account'', in S. Shoemaker and R. Swinburne (above n. 4) and elsewhere, Shoemaker offers an argument from functionalism in the philosophy of mind to the Wide Psychological View. This argument is well met by Eric Olson in ''What Does Functionalism Tell Us About Personal Identity?'' *Nous* 36 (2002). See also my review of the Shoemaker and Swinburne volume in *The Philosophical Review* 96 (1987).

''personas'', that is, the psychological and social continuers of developed persons. In saying that, we need not give up on the ordinary thought that persons predate and outlive their personas.

The third potentially distorting effect on our intuitions mentioned in ''Human Beings'' is a misleading reification of an unspecific concept of ourselves as ''this subject of experience''. That way of picking ourselves out, associated with (but not semantically equivalent to) some uses of ''I'', does not make any commitment as to the specific nature of the subject picked out. If one mistakes this lack of specificity in the way in which one is thinking of oneself in first-person thought for a lack of specificity in the thing we are thinking of, or simply elide that distinction, then one can readily imagine surviving any amount of psychological and bodily discontinuity, as if one were a pure ego or Bare Locus of mental life. Thus I can imagine, from the inside as it were, turning into a dolphin, that is, undergoing a wild transformation in my body and psychology, thinking all the while ''What is happening to *me*?''

So the method of cases stands seriously impeached, as does the Wide Psychological View, which appears to depend on an appeal to intuitions about *fringe* cases.

4

What then is the alternative methodology for theorizing about personal identity; equivalently, about what kind of thing we are? At one level, it can be none other than the method of real definition; namely using all of the relevant knowledge and argumentative ingenuity we can muster in order to say *what is it to be* the given item or phenomenon. ''Human Beings'' made a more concrete proposal as to how to approach such a real definition. The humble and ubiquitous activity of tracing ourselves and others over time places significant constraints on the kind of thing we are. Whatever kind it is, you won't get very far saying that it's a kind whose members we would find it very hard to trace in the easy and offhand ways in which we seem to trace ourselves and others. For the cost of saying that comes close to the severing of all connection with the topic of personal identity.

So, for example, we can set aside the idea that we are Bare Loci of mental life, barely constrained either by bodily or by mental continuity. Perhaps the idea that we are separable mental substances that are only

remotely connected to bodily activity also can be set aside. The argument is not that these things do not exist, although I do reserve the right to argue that. The argument is that even if they do, they are not the things that we are tracing when we trace persons. Otherwise, our offhand methods of tracing by way of carefree observation of bodily continuity, sometimes augmented by observing some sort of consistency in behavior, would be woefully inadequate to the task of tracing persons. Relative to our easy and offhand methods of tracing persons, judgments about the persistence of Bare Loci or separable mental substances would be highly adventurous conjectures. And yet we do have knowledge of personal identity over time, at least in the massive core of ordinary cases. Indeed, it is among the most secure sort of knowledge that we possess.

So the suggestion was that our point of departure in theorizing about personal identity should be the hypothesis that we are some natural kind that is ''epistemologically ready to hand'', that is, one whose members are easily traceable in the undemanding ways in which we trace ourselves and others. A good place to start, I said, was with the biological kind *Homo sapiens sapiens*, and with the concomitant thought that we are certain kinds of animals. That point could be reinforced by observing that much of our tracing of our fellows through time is strongly akin to the tracing of cats by dogs, or dogs by dogs, or dogs by us. Similar perceptual and cognitive mechanisms appear to be involved. And in all these cases, the object of attention is (almost) unquestionably another animal.

5

Now it has been the position of those who style themselves ''Animalists'', most notably Paul Snowdon and Eric Olson, that this point of departure is also the terminus of our journey in theorizing about personal identity.[6] For them, the answer to the question with which

[6] Paul Snowdon ''Persons, Animals, and Ourselves'', in Christopher Gill (ed.), *The Person and The Human Mind* (Oxford University Press, 1990); Eric Olson, *The Human Animal* (Oxford University Press, 1997). I think of these writers as reviving the position of David Wiggins, who wrote ''a person is any animal that is such by its kind to have the biological capacities to enjoy fully the psychological capacities enumerated''; and, ''There would be no one real essence of persons as such, but every person could still have the real essence of a certain kind of animal. Indirectly this would be the real essence in virtue of which he was a person'' (Wiggins, *Sameness and Substance* (Oxford University Press, 1980), 176).

we began is that we are essentially members of the biological kind *Homo sapiens sapiens*. And I know from my very pleasant conversations with him, that Snowdon, at least, regards my resistance to this obvious position as particularly puzzling. For, after all, the objection that the Animalists regard as putting most pressure on their position is one that can be dismissed given the resources developed in "Human Beings"!

Let me explain. Thomas Mann, during his wartime sojourn in Princeton, New Jersey, wrote a novella entitled *Transposed Heads* (1945), in which the fantastic conceit of head-swapping is given a brilliant seriocomic treatment. (It is worth looking into Mann's implied judgments about who turns out to be whom!) It seems that the credit for introducing a roughly equivalent conceit into philosophical discourse should go to Sydney Shoemaker, who writes in his *Self-Knowledge and Self-Identity*,

> It is now possible to transplant certain organs . . . in such a way that the organ continues to function in its new setting. . . . [I]t is at least conceivable . . . that a human body could continue to function normally if its brain were replaced by one taken from another human body. . . . Two men, a Mr. Brown and a Mr. Robinson, had been operated on for brain tumors, and brain extractions had been performed on both of them. At the end of the operations, however, the assistant inadvertently put Brown's brain in Robinson's head, and Robinson's brain in Brown's head. One of these men immediately dies, but the other, the one with Robinson's head and Brown's brain, eventually regains consciousness. Let us call the latter "Brownson." . . . When asked his name he automatically replies "Brown." He recognizes Brown's wife and family . . . and is able to describe in detail events in Brown's life . . . . Of Robinson's past life he evidences no knowledge at all.[7]

As a matter of statistical fact, by far the most common reaction to this case is to suppose that Brownson is Brown. Yet it seems clear that Brownson is not the same organism or human animal as Brown. In fact, Brown's de-brained body—let's call it "Brownless"—could be provided with enough in the way of brain stem tissue, transplanted from still another source, so as to be kept alive in what has come to be called a persistent vegetative condition. It can seem that this human "vegetable" Brownless is identical with the very human *animal* that

[7] *Self-Knowledge and Self-Identity* (Cornell University Press, 1963), 23–4.

once exhausted Brown's bodily nature. Brownless is that animal in a surgically mutilated condition.

If that is so, and if Brown is not identical with Brownless, but with Brownson, then Brown and his kind-mates, those essentially the same as he, are not essentially human animals. If, in such cases, you go where your brain goes, then you are not essentially constituted by a human organism, even if in all ordinary situations you are wholly constituted by one.

Note that this is a classic appeal to an intuition about a fringe case, a case which is *precisely designed* to tease apart mental continuity (holding between Brown and Brownson) and the organic physical continuity (holding between Brown and Brownless) that secures the survival of a human animal. As Snowdon urged, shouldn't the author of ''Human Beings'' be the first to reject the appeal to such cases, on the grounds that in separating out mental and physical continuity they may well undermine the contingent preconditions of our being good judges of personal identity?

Furthermore, the standard reaction to this case can seem to have the flavor of a sort of ''Materialist Cartesianism''. Here is the brain, kept alive and functioning during the transplant. It is easy to imagine a self contingently attached to it, a self understood as a Bare Locus of mental life, or as a separable mental substance. In tracing Brown so that he ends up as Brownson, might we not be misled by imaging the brain as a sort of *organic chariot* for such a Bare Locus, or such a separable mental substance? If that is a likely possibility, if we are likely to be thinking of the brain as a chariot of the pure ego, shouldn't the author of ''Human Beings'' be the last person to credit the dominant intuitive reaction in such cases?

Moreover, notice that Shoemaker neatly frames his question about identity in precisely the terms that would invoke the psychological and social continuer effects. He writes:

> When asked his name [Brownson] automatically replies ''Brown.'' He recognizes Brown's wife and family . . . and is able to describe in detail events in Brown's life. . . . Of Robinson's past life he evidences no knowledge at all.

Shouldn't the author of ''Human Beings'' be most suspicious about this sort of framing of an imaginary case? So might my friends the Animalists berate me!

But of course, I agree with them on all three counts, and said as much twenty years ago. I wrote:

Usually . . . the ''predominant'' reaction to Shoemaker's case, *viz*., the fact that most judge that Brownson is Brown, is taken as establishing that Brownson is Brown. But this is just another example of uncritical reliance on the method of cases. Indeed, the predominant reaction can be explained in terms of the distorting influences described below as the psychological- and social-continuer effects. *We* need a principled reason, consonant with the alternative method, for taking seriously the judgment that Brownson is Brown and thereby departing from the straightforward naturalistic view of ourselves as essentially human organisms.

It is noteworthy that the most forceful defender of Animalism, Eric Olson, writing ten years later, appeals to what amounts to the psychological and social continuer effects, precisely to discredit the intuition that you go where your brain goes.[8]

6

So why am I not an Animalist? It is because Animalists are badly placed to account for the status of ''remnant'' persons.

Suppose Brown's whole head and neck is removed, and kept alive and functioning by artificial means. His torso, or ''head-and-neck complement'' is destroyed. Brown's head is not an animal. An animal is wholly constituted by an organism, something with the power of self-maintenance in the natural environment in which the animal's species developed. The head and neck can be maintained as a functioning unit only by extrinsic, highly artificial means. Even if we consider that the larger whole includes the head, the machines, and the medical staff, we do not find anything that deserves the name of ''an animal''. So, in holding that ordinary human persons are essentially animals, Animalists are committed to the claim that Brown does not survive as his separated head and neck.

This is philosophy, so we should be prepared for the comeback that the head and neck *is* an animal, perhaps because it includes what was a controlling part of the life of an animal, namely the brain. But then consider the head without the neck; it too includes the brain, so is it an animal? But then consider the brain, and then the cerebrum, on the assumption that something like the cerebrum is the minimal part that

[8] Eric Olson, (above n. 6) 68–70.

could subserve reflective mental life of some complexity. Is the brain an animal? Is the cerebrum an animal? The argument that follows could be run through with any such minimal remnant. Do we want to say that any such minimal remnant is an animal? Surely not. An animal is (constituted by) an organism, not by a mere organ or mere organ part.

The crucial point to notice is this: In order to disable the following argument from remnants, the Animalist would have to argue that *any* remnant that is complex enough to be a locus of reflective mental life counts as an animal. And that would depart significantly from any ordinary sense of ''animal''. For that reason, Animalists have supposed that such remnants are not animals.

So consider such a remnant, Brown's separated head and neck, kept alive and functioning. Suppose it is provided with a continuous flow of air through its larynx and mouth. We might then observe a lengthy diatribe emanating from Brown's head. In this, and other ways, we might come to have massive evidence that Brown's head, or the head and its mechanical sustainers, constitutes a person in John Locke's sense of ''a thinking intelligent being, that has reason and reflection, that can consider itself as itself, the same thinking thing, at different times and places''. Let's call the ostensible person constituted by Brown's head, or the head and its mechanical sustainers, a ''remnant person''.

Animalists are happy to use John Locke's characterization of a person. Their point is simply that we human persons turn out to be animals. But what of remnant persons? Animalists need not implausibly say that these would be animals too. But they will then have either to deny that there would be such remnant persons or maintain that they come into being with the destruction of the ''head-and-neck complements'' of human animals.

As to the first option, it looks like a mistaken philosophical strategy. For it involves foreclosing something that any philosophical theory should leave as an open empirical possibility. It is likely that we could have massive evidence that there was such person in the vicinity of a severed head, if it were kept alive and functioning. *He* could vociferously lament his condition; or describe pains in his phantom body, or complain about the paralysis of his phantom limbs.

The remaining option is to suggest that a remnant person would come into being as a result of the destruction of the torso or head-and-neck complement of a human animal. The Animalists should say that the

person constituted by Brown's head and neck could not be Brown. But they can allow that it is a new, numerically distinct, person who was produced by the destruction of Brown's head-and-neck complement.

Now here is a general conviction that many will share, one which organizes some of our thinking about persons and physical reality: You can't bring a person into being simply by removing tissue from something, and then destroying that tissue, unless that tissue was functioning to suppress mental life or the capacity for mental life. A developing fetus might have a massive tumor in its developing brain, which suppresses its mental life, and perhaps even its capacity for mental life. Given that, we can understand how removing the tumor could allow a person in Locke's sense to be present for the first time. But how could removing a sustaining torso bring this about? The presence of a person seems not to depend on this sort of extrinsic matter. (I cannot see how, in accepting this last principle, we must be implicitly relying on some argumentative route that is prey to any of the distorting influences detailed in Section 5.)

Thus, by denying that Brown survives as the person constituted by his separated head and neck, the Animalists run out of good options when in comes to accounting for remnant persons. They must either deny that there could be remnant persons or suppose that we bring such persons into being by removing and destroying tissue that was not functioning to suppress mental life, or the capacity for mental life.

7

The important thing to recognize is that in thus rejecting Animalism, one need not reject the plausible naturalistic view that we are animals, and that in tracing ourselves and each other in the massive core of ordinary cases, we are tracing things which are animals. We may distinguish two claims, accepting the first but not the second.

> We are animals, members of the species *Homo sapiens sapiens*.
>
> We are essentially animals, and so such that we are animals in every possible situation in which we survive.

The Animalist's strongest argument, the "too many minds" objection, only supports the first claim. Yet Animalists like Eric Olson write as if it went all the way to supporting the second.

In ''An Argument for Animalism'' Olson argues that we are animals as follows:

I turn now to my case for Animalism. It seems evident that there is a human animal intimately related to you. It is the one located where you are, the one we point to when we point to you, the one sitting in your chair. It seems equally evident that human animals can think. They can act. They can be aware of themselves and the world. Those with mature nervous systems in good working order can anyway. So there is a thinking, acting human animal sitting where you are now. But you think and act. You are the thinking being sitting in your chair.

It follows from these apparently trite observations that you are an animal. . . . and there is nothing special about you: we are all animals. If anyone suspects a trick, here is the argument's logical form.

1. (∃x) (x is a human animal & x is sitting in your chair)
2. (x) ((x is a human animal & x is sitting in your chair) → x is thinking)
3. (x) (x is thinking & x is sitting in your chair → x = you)
4. (∃x) (x is a human animal & x = you)[9]

The motivating thought is that animals, especially highly evolved human animals, can think; that is, can be subjects of belief, desire, feeling, and sensory experience. But if you are not identical with such an animal then there are, bizarrely, at least two thinkers occupying your chair. For there is an animal there, and you are, indisputably, a thinker.

Olson supposes that resistance to his argument will take one of three forms: either there are no human animals, or human animals can't think, or there are indeed two thinkers there. Speaking for myself, all of these options seem a little silly. (There may be other reasons, arising from cases of commissurotomy and the like, that could prompt us to multiply thinkers justly, but it would be absurd to do this just because of the ordinary case of a person sitting in his chair.)

I think Olson has a sound argument here, one that it is a little silly to reject, and one that must condition all further discussion of personal identity. *We are animals.*

[9] In J. Barresi and R. Martin (eds.), *Personal Identity* (Blackwell, 2003). The argument is taken from pp. 325–6.

8

Elsewhere, I have taken some trouble to argue that constitution is not identity.[10] There is, however, a mistaken deployment of that thought which is very seductive in the present context. It is important to see why it is mistaken.

Shouldn't an advocate of the view that we are essentially human beings say that although a human being is *constituted* by a human animal it could also be constituted by a head (or brain, or head and neck) kept alive and functioning? If so, the diagnosis of Olson's master argument would then proceed as follows:

> The first premise has a reading on which it is true. Yes, there is something *constituted* by a human animal that is in your chair. The second premise has a reading on which it is true. Yes, everything constituted by a human animal and that is also in your chair, namely you, is thinking. The third premise is straightforwardly true: you are the only thing in your chair that is thinking. So the conclusion should be now read as: There is something which is constituted by a human animal which is you. But that constituted thing is a human being.

The problem with this diagnosis is that it is something of a platitude that higher *animals* can ''think'', that is, at the very least they can feel, and sense their environment. Non-human animals do this all the time. (Some philosophers have denied this, I am sad to report.) Why should highly evolved animals like us, with our highly complex brains, be unable to do this? If animals can feel and sense their environments, why should some animals fail to do this just because they wholly constitute something else? Would you fail to be a thinker while you were working as, and so constituting, a living street-statue of Dante? No, the animal that allegedly constitutes you is now seeing the scene before it, and you are too. If you are not that animal, this is a bad result; we have one too many things seeing the scene before you.

Some are tempted to say that a human person counts as seeing the scene before him *because* the associated animal does. We have two

[10] ''Constitution is Not Identity'', *Mind* 101 (1991), reprinted in Michael Rea (ed.), *Material Constitution* (Cornell University Press, 1995). See also ''Constitution'', in F. Jackson and M. Smith (eds.), *Oxford Handbook of Philosophy* (Oxford University Press, 2005).

thinkers, the person and the animal, but the person counts as a thinker derivatively. In saying this they may have in mind the analogy with weight; a statue is not identical with the bronze that makes it up. But if the bronze weighs ten pounds, then the statue does too; for a material item's weight is just the weight of its matter. The statue ''inherits'' the property of weighing ten pounds from the matter that constitutes it. Once you recognize this, you will feel that there is nothing paradoxical about the thought that in a given region there are two ''things'', the statue and the bronze, that each weigh ten pounds. There is for example no temptation to conclude that the weight of the things in the region occupied by the statue is twenty pounds. Once we accept this, doesn't the ''too many thinkers'' objection look to be something of a trick, which we already know how to unravel?

The crucial point, the point that changes one's take on the ''too many thinkers'' objection, so that it will no longer seem like a mere trick to the friends of constitution, is that the analogy with weight gives the wrong result! It is not a good way out to say that the property of thinking is ''inherited'' by the person from the animal that constitutes it. For if I had to pick which of the two things I am identical with, the person or the animal, a good rule would be: Pick the thing which is *non-derivatively* the subject of mental acts. And on the present proposal with its appeal to ''inheritance'' it is the animal, and not the person, that is non-derivatively the subject of my mental acts. I would be identical with the animal, and so not with the person it constitutes. But the result we wanted was that I am identical with the person. That is a datum which any theory must capture.

So claiming that a human being or a human person is constituted by a human *animal* simply plays into Olson's hands. For we can't plausibly deny that an animal thinks, that is, senses and feels. This is why we should accept Olson's argument as he intends it. I am not *constituted* by a human animal. I am (this is the ''am'' of predication) a human animal.

What I reject, however, is the next consequence Olson draws from this argument. Speaking of the position which he believes it supports, he writes:

> But what of the case against Animalism? It seems that you would go along with your cerebrum if that organ were transplanted. . . . And that is not true of any animal. Generations of philosophers have found this argument compelling. How can they have gone so badly wrong?[11]

[11] Eric Olson, (above n. 6), 331.

Olson's answer is that they have mistaken a certain sort of evidence for personal identity, namely psychological continuity, for what personal identity consists in. (That is the thing I would say myself, if I held his view.)

However, as noted earlier, Olson's view has the repugnant consequence that we can make a new person simply by removing tissue which previously played no role in suppressing mental life, or the capacity for mental life. Or it has the repugnant consequence that there could be no person associated with my severed head, even when that head is kept alive and functioning, when the rest of my original body has been destroyed.

What has gone ''so badly'' wrong? An unstated assumption has slipped into the argument.

9

Why should the conclusion of Olson's master argument (line 4) be taken to be at odds with the claim that you would go where your brain (or cerebrum) goes?

Consider this argument, directed at an adolescent sitting in her chair.

5. (∃x) (x is an adolescent & x is sitting in your chair)
6. (x) ((x is an adolescent & x is sitting in your chair) → x is thinking)
7. (x) (x is thinking & x is sitting in your chair → x = you)
8. (∃x) (x is an adolescent & x = you)

Again, this is a perfectly sound argument. Yet the conclusion is obviously consistent with the fact that the adolescent can survive the process of ceasing to be an adolescent and becoming an adult. As philosophical logicians used to say, ''adolescent'' is a phase sortal. It applies to a thing only during a certain phase of that thing's existence. What isn't true is:

9. If x is an adolescent then x is essentially an adolescent.

Along with phase sortals, that is, kind classifications which apply temporarily to things that they classify, there are contingent sortals, that is, kind classifications which happen to apply to things throughout their entire careers, but need not. By contrast with contingent sortals and phase sortals, there are ''substance sortals'', which classify things by their essential features.

Animalists may have implicitly assumed:

10. ''Animal'' is always a substance sortal; that is, if anything is an animal then it is essentially so.

Once articulated, this looks like quite an embarrassing assumption to have made in the context of a discussion of brain (or cerebrum) transplants. For, the Animalists' opponents, some of them at least, are suggesting that during the transplant, while the brain (or cerebrum) is kept alive and functioning the original person has been reduced to the condition of a brain (or cerebrum). And that view could be put this way: before the transplant the original person was a human animal, and during the transplant he is not an animal. (No brain or cerebrum, and indeed no head, constitutes an animal.) This entails that

11. ''Animal'' is not always a substance sortal; there are some things that are contingently animals.

Indeed, my somewhat regimented use, twenty years ago, of the term ''human being'' was intended to pick out just such a thing. So long as we do not mistakenly endorse 10, the alternative view that we are essentially human beings—namely, things that in every ordinary case are human animals, but which could be reduced to the condition of a mere head or brain—is entirely consistent with the real insight in Animalism, namely that we are animals.

10

Animalists initially may resist the idea that there is some natural kind classification like ''animal'' which applies essentially to some class of things, say worms, and contingently to another class of things, say human beings. Nevertheless, this is something that Animalists need to get used to anyway.

For one example, consider the charge e on the electron. This looks as though it is had essentially by the electron; physicists will not call anything that lacks the charge e an electron. But there could be variable composites of particles whose charges at some time sum to e. Such a variable composite could lose one of its constituents, and so come to have a different charge from e. So, the sortal ''physical item with charge e'' is a substance sortal when applied to electrons, and a phase sortal when applied to such variable composites of particles.

Among the things classified by natural classifications are species themselves, changing items in space and time whose varying constitution consists in different populations of animals or plants at different times. So far as I can see, it might well be that thanks to their simple digestive systems, certain animals that are herbivores more or less must be so; while others that are actually herbivores could come to be carnivores or omnivores. Here again, we have the possibility of a natural classification that holds essentially in some cases and contingently in other cases.

For another example, consider ploidy, a measure of the number of copies of its basic chromosomes that an organic thing produces. The ploidy of cells can vary within an organism. In humans, most cells are diploid, containing two copies of the chromosomes, though sex cells are haploid, containing a single copy of each chromosome. The ploidy of a simple cell *might* be held to be essential to it, so that for a cell, ''haploid'' and ''diploid'' are substance sortals. But a multi-cellular organism might alternate between grossly haploid and diploid conditions. It seems that algae are in fact like this.[12]

These considerations might be combined with a rather humdrum speculation: after my conception it took me some time to develop into an animal. For a while, I was a multi-celled organism—a pre-fetal embryo—wholly parasitic on my mother. I developed into an animal, and was not always an animal. So although I am (now) an animal, I once was not.

Earlier, it was said that such speculations are not probative; there are many things we could say about the beginnings of life. It is one thing to admit that, and quite another to say that I *must* always have been an animal because I am one now! It is clearly a substantial, even controversial, assumption that if a natural classification like ''animal'' applies essentially to one class of things, it applies essentially to all the classes of things to which it applies.

There is then room for the claim that we human animals are human beings, who *could in principle* be reduced to the condition of a head, or of a brain, kept alive and functioning, even though no animal could survive in this condition. We are however, for all practical purposes locked into a ''permanent'' phase, that of being an animal. ''Human animal'' is what we might call a ''crypto-phase sortal''. It is not an

[12] Thanks to Karen Bennett for drawing my attention to this phenomenon.

obvious phase sortal like ''adolescent'', but it can be shown to be a phase sortal upon philosophical reflection.

11

The ''too many minds'' objection may now seem to have been based on a simple mistake about sortal predication. To return to Locke's famous characterization of a person, no one would deny that an adolescent is ''a thinking intelligent being, that has reason and reflection, that can consider itself as itself, the same thinking thing, at different times and places''. (Well, no one other than the adolescent's parents would deny this.) But then there is the person who is the adolescent, and can cease to be an adolescent. So do we have two persons in the same place, at the same time? Of course not, the person is (predicatively) the adolescent; there is one thing that is a person and an adolescent. So also, there is one thing here which is a human being, an animal, a person, and an adolescent. Only the first of these classifications bears on the essence of the thing in question.

That might seem reassuring, but I think that it is an incomplete response to Olson's ''too many minds'' objection. Despite its disarming simplicity, it is a subtle objection, which Olson has elaborated in convincing ways, and it needs to be handled with some delicacy. To meet that objection fully we need to respect the idea that the things that have minds, the things that ''think'', are persons or animals. (Leave thinking machines out of it for the moment.) So much follows from the logical grammar of our predications of mentality. If we are saying that something is seeing or thinking or feeling then the subject of predication should be an animal or a person. I do not say that such logical grammar is ontologically very deep, but only that it organizes our thought about thinking things, and that if we fail to respect it, we will be led astray.

Nor do I say that good philosophers always respect it. Just to take one notable example, Patricia Churchland, in her very fine book *Brain-Wise*, is continually saying that *brains* see, hear, plan, imagine, remember, and so on.[13] Of course I know a paraphrase on which what she says comes out true. Still, remarks like ''the brain sees'' make me balk, and largely because of the considerations that Olson has so well articulated.

[13] *Brain-Wise* (MIT Press, 2002), *passim*.

You are now seeing the scene before you, and you are not identical with your brain. You are bigger than your brain. And it is just lazy to suppose that you *derivatively* count as seeing the scene before you *because* your brain does. For, if one had to pick which thing one is, a good rule would be: pick the thing which is non-derivatively the source of your mental acts. So you are your brain, you have been encased in your head all this time! That is an absurdity that follows from the idea that the brain sees or thinks. The same absurdity can be derived from the idea that a head sees or thinks.

Could a body or a human organism be said to think? Obviously ''thinking human organism'' and ''thinking human body'' are weird phrases; but the *argument* that your body or organism cannot think only emerges if we suppose (rightly on my view) that you, the person, are wholly constituted by, but not identical with, your body. Then, if your body thinks, there will be two thinkers, the person you are and your body; and it looks as though your body thinks in a more primary way. Thus we have it that you are your body, contradicting our supposition that you are the person constituted by that body.

So we should deny that a brain or a head thinks, and also deny that an organism thinks, in any primary, non-derivative sense. Instead we should say that these are things such that some of their operations *constitute* the thinking of a person or animal. (That might be loosely described as ''thinking'', just as some natural structure might be called a shelter.) Our concept of a thinking thing is non-accidentally and primarily tied to *persons* and *animals*.

But, it will be objected, isn't this true:

(x)(x is a animal → x is an organism)?

And if it is possible for an animal to think, surely it follows that it is possible for an organism to think. By this route, doesn't the ''too many thinkers'' problem come back to haunt us, after all?

No. What is true is:

(x)(x is an animal → x is *wholly constituted by* an organism).

I am an animal. This organism here, the H. M. S. Johnston, is my body. *But my body is not an animal.* Therefore, I am not identical with my body. Therefore, the animal that I am is not identical with my body. ''My body'' denotes the organism that constitutes the animal that I am. Some of the operations of that organism constitute my thinking. They

provide the matter of my thoughts, in a way analogous to that in which my body provides my matter.

I have been in odd discussions, where people talked indifferently of ''my body'' and ''my animal''. Yet it must be said that the latter is a very odd usage; unless, say, I'm talking about my dog. I am wholly constituted by an organism, I have a body, and I *am* (the ''am'' of predication) an animal. ''My animal'' is too close to me to be *mine*. It is *me*. To speak of my animal is like speaking of *my me*. As with phrases like ''*Das Ich*'', you might be able to give such phrases as ''my animal'' a use in a philosophical discussion, but you should be very clear about what it takes to do that.

In contrast, there is nothing wrong with the phrase ''rational animal''. I am a rational animal, even though as a real definition ''I am a rational animal'' is a failure. But ''rational organism'' and ''rational body'' are just weird. Ordinary logical grammar seems to be signaling that animals and organisms are not one and the same.

If we respect this ordinary logical grammar which frames talk of one's body, the organism that one is constituted by, and the animal that one is, then we will see that the ''Organismic'' argument that models itself on the sound Animalist argument supports a different sort of conclusion:

12. $(\exists x)$ (x is a human organism & x is in your chair)
13. (x) ((x is a human organism & x is in your chair) $\rightarrow$ x is thinking)
14. (x) (x is thinking & x is in your chair $\rightarrow$ x $=$ you)
15. $(\exists x)$ (x is a human organism & x $=$ you)

The first premise has a reading on which it is true. Yes, there is something *constituted* by a human organism that is in your chair. That is the only sense in which you are an organism. The second premise has a reading on which it is true. Yes, everything constituted by a human organism and that is also in you chair, namely you, happens to be thinking. The third premise is straightforwardly true: you are the only one in your chair that is thinking. So the conclusion should be now read as: there is something, you, which is constituted by a human organism. A truth no doubt, *but a truth that now has been immunized against the virus of ''too may minds''*. While it is absurd to deny that a healthy animal with a complex functioning brain can think, we have independent grounds to deny that a brain or a head or an organism can think.

The suggestion that organisms (bodies) constitute animals can be filled out by describing a situation where we have the organism (body) that constituted an animal, but not the animal itself. It seems to me that there are such situations; death is the end of an animal's life, and hence the end of an animal. But it need not be the end of the organism that constituted the animal. A dead organism is still an organism. A dead body is still a body. When we talk of "dead animals", we are talking of dead organisms or dead bodies that constituted animals.

I suppose the Animalist might argue as follows: just as a body is contingently alive, an animal is also contingently alive. A dead parrot is not an ex-parrot, contrary to Monty Python. It is still a parrot, it is just a parrot that is dead. But then I would say (putting aside all hopes of an afterlife, at least for now), surely death is the cessation of *my* existence, a dead person is an ex-person. That can make things look bad for the Animalist. For he wants to claim that there is something that is me and also an animal, and he wants to claim that "animal" is here functioning as substance sortal. Yet the animal in question exists for a while after my death, until it decays. I do not exist for a while after my death.

I suppose an Animalist might deny that, say by construing our talk of recently dead people as talk of existent people who happen to be dead. One consequence of this is that it could be literally true that Paddy died, and *then*, as if that had not been bad enough, a week later at his cremation, he *ceased to exist*!

I offer the Animalists an alternative, one which sides with Animalism *in insisting that we are animals, members of the species* Homo sapiens sapiens, but avoids strange commitments, such as the claim that we can make a person merely by tissue-removal, even though the removed tissue was in no way suppressing mental life or the capacity for mental life.

The alternative goes like this. You are a human being. Like many things that persist through time, you have a variable constitution over time. Perhaps you were first constituted by a multi-cellular organism, a pre-fetal embryo, and this organism turned into a living human body. As to whether that body is the original organism in a highly developed condition, and so is just a phase of a longer-lived organism, we may remain agnostic. In any case, without settling that question, we can say that you were first constituted by an organism, then by a body, that is, an organism of the sort that makes up an animal. Moreover, in the bizarre event of your head-separation, and the destruction of the rest of

your body, you would then come to be constituted by a head, something which is neither a human body nor an organism. You would cease to be an animal, but you would not cease to be.

You are an animal at all those times at which you are wholly constituted by the right sort of living organism. Plants are organisms too, as are zygotes; they are obviously the wrong sorts of organisms. Perhaps "body" is a good name for the right sort of organism. As long as you *have* a body, and it is alive, you are an animal. The very beginnings of our lives aside, in all real cases we have ever encountered, we are animals.

12

Thus, the sound Animalist argument to the conclusion that we are, all of us, now animals is thus entirely compatible with the view defended in "Human Beings". It *is* at odds with the Wide Psychological View, at least when that view is understood as defining a person as a cross-time bundle of mental events and states united by relations of psychological connectedness and continuity. (The Wide Psychological View can satisfy materialist scruples by requiring further that each mental state and event in the bundle is physically realized, though not necessarily by the same animal, brain, or human being.) Such a bundle of mental events and states is not predicatively an animal at any time. For no animal is such a bundle of mental events and states, or a slice of such a bundle. And yet, you and I are now animals. When we trace each other in the ordinary ways we do, we are tracing things that are animals. So the view that treats us as such cross-time mental bundles is not to be favored, once we adopt the methodology outlined in "Human Beings".

Since a person is a thinker, the friend of the Wide Psychological View, who treats a person as a bundle of mental events and states, had better say that such a bundle can think. Then such a theorist faces a version of Olson's problem: he must hold either that there are no human animals, or that human animals can't think, or that there are indeed two thinkers in all the ordinary situations where we are convinced there is just one. And aren't these options rather silly, or at least very unattractive?

The foregoing shows that the Wide Psychological View had better bundle up something which is, under all normal conditions, predicatively an animal. This may seem a simple matter. Why not bundle in with the

mental events and states the very physical events and states that realize them? The cross-time unity conditions for the bundle, the gen-identity conditions in the sense introduced earlier, will still be purely psychological, and still hold among mental events and states. But the physical realizers of these states and events, namely physical events and states of the bodies of animals, will also be included in the bundle. There is now no straightforward argument that neither the bundles nor their time slices can be said to be animals.

Still, I shall argue that when we trace ourselves and others through time we are not in fact tracing such bundles, at least if nature has been wise enough to have us rely on the method of offloading. This will be part of the burden of my response to Denis Robinson.

13

In his insightful and probing discussion of ''Human Beings'' (chapter 1, this volume), Robinson, who favors the Wide Psychological View, is pleased to goad me with the very good question of whether, given my alternative methodology, I have any principled reason to reject Animalism. In so doing, Robinson hopes to engage me in an unpromising battle on two fronts. Alas, such is the fate of moderate compromises between extreme positions!

Perhaps enough has been said about Animalism, but I should note that Robinson's systematic substitution of the term ''human animal'' for my preferred term ''human organism'' is not as harmless as he seems to imagine. It leads him to mischaracterize what on my view counts as a potentially temporary condition of human beings, and what typically constitutes human beings. In ''Human Beings'' I was concerned mostly with the latter, and I took some stylistic trouble to systematically talk of ''human *organisms*'' as the constitutors of human beings. Robinson ignores this, and substitutes the phrase ''human animals''. So contrary to the impression that Robinson gives, it was not, and is not, my view that human beings are constituted by animals! My body is not an animal, although I am an animal, thanks to being wholly constituted by that body, or human organism.

Let us turn to the Wide Psychological View, which has it that psychological continuity (holding in a one–one fashion) is sufficient for the

survival of a person, even if that continuity is secured by a series of distinct brains or brain-like-items.

Recall that the alternative methodology began by asking: What kinds of things are such that they could be traced in the easy and offhand ways in which we trace ourselves and each other over time? Call the kinds that pass this test ''the initial pool''. We have been looking in detail at two candidates in the initial pool, the kind animal (understood as applying to us essentially) and the kind human being. Even if the kind human being fares the better of the two, are there not other candidates in the initial pool? In this vein, Robinson writes:

> A central weakness of Johnston (1987) is the lack of any overt argument in favor of the preference, albeit a fashionable one, for a pool which contains biological and other natural kinds, but which excludes mentalistically individuated substances per se. Johnston's choice here (to allow only non-mentalistically individuated entities as bona fide substances) is not altogether arbitrary or unmotivated. He claims that the method of cases fails in this context, in part because ''the'' concept of a person, taken as what is in common to all the different conceptions of personhood which have historically been held or which may emerge from reflection on imaginary cases and the like, is impossibly vacuous. Nothing but a soul pellet could fit it, yet a capacity unproblematically to reidentify soul pellets would be impossible to account for; nor would it make sense to attach the kind of importance to the vicissitudes of soul pellets which we in fact attach to issues of personal identity. Thus it is natural for him to assume that his alternative method must to some extent rely on *nature*, rather than on *our concepts*, to demarcate the persons: to assume, in other words, that the kind ''Human Person'' must be constructible relatively straightforwardly out of natural kinds. And in this context, as noted, Johnston identifies the relevant natural kinds as the biological kinds. (Ch. 1, pp. 17–18)

Robinson goes on to register the fact that I myself indicated that more needed to be said to remove the Wide Psychological View from contention, and that this would take a detailed discussion of the persistence of various kinds of things.[14] (Some of that will be provided below.) In any case, he asks another very good question: In tracing ourselves and others in the easy and offhand ways in which we do, how can we be so sure that we are not tracing ''mentalistically individuated'' items?

[14] See Ch. 1, n. 19.

Perhaps because he favors the Wide Psychological View, and thinks of that as the main contender, Robinson does not provide other examples of what he calls ''mentalistically individuated'' items. So let's examine a few, first in the category of substances, and then in the category of cross-time bundles.

14

Are we Bare Loci of mental life, ''soul pellets'' that could survive any amount of psychological or physical discontinuity? One might think of a Bare Locus as a reified mental haecceity, something that carries the identity of a person, and can attach to this or that body or functional psychology. In one way, this conception fits with the wilder possibilities of reincarnation, as when I am supposed to be able to come back as a frog. In another way, it does not. Why should a Bare Locus bear any Karmic responsibility for the mental and physical life to which it happened to be attached? The Bare Locus is ontologically too thin a thing itself to make any choices that warrant Karmic retribution.

The argument of ''Human Beings'' was that we are not Bare Loci, because these are not the sorts of things that could be traced in the easy and offhand ways in which we trace ourselves and each other.

Are we then ''Cartesian'' mental substances, primary bearers of mental properties, which are not dependent for their survival on the survival of the organism or the brain with which they causally interact? This view has a variant on which people, ''the men and women themselves'', are not identical with such mental substances. They are composites of bodies and mental substances, composites whose principle of unity is that the mental substances do interact with the bodies in question. So on this variant, when the body dies, the man or woman in question ceases to be, even though his or her soul or mental substance continues on.

Either variant of the ''Cartesian'' view would be made quite probable if we took at face value the reports that appear in roughly ten percent of all successful resuscitations after cardiac arrest. These are reports of leaving one's body, while still being a thinking and perceiving thing, indeed a thing that occupies a point of view above one's body, and the doctors trying to resuscitate it.

No account of our essence should ignore such evidence, *if it really is evidence*. However, the alternative method outlined in ''Human Beings'' still constrains what we should say if we *were* persuaded by the problematic, post-resuscitation reports. The things that we have been tracing, in the easy and offhand ways in which we trace ourselves and each other, are more likely to be the composites rather than the souls, for it is the composites that have the manifest bodily configuration that captures our attention. Our tracing is not a highly inferential matter, as it would have to have been if we are to be credited with having ever traced Cartesian mental substances. So the first viable position to consider is the Cartesian Composite View, on which a person is a composite.

I do not want to enter my skeptical responses to the post-resuscitation reports here. Our question should rather be: Is there a philosophical argument against the Cartesian Composite View? Well, for one thing, it is not clear to me that the composite couldn't be called ''an animal''. Accepting the view might be held to be accepting a novel account of what the members of the species *Homo sapiens sapiens* consist in, namely a soul and a body. The view would in no way deny the bodily, organic aspect of our natures. After all, the composite would cease to exist when the animal dies. Only the mental substance or soul lives on.

The argument from tracing is less conclusive against the Cartesian Composite View than it was against the view that we are Bare Loci. For very young children readily understand tales about princes turning into frogs, so long as the princely mentality is carried across into the froggy body. Perhaps in tracing persons as opposed to mere bodies, children are deploying an implicit conception of a person as an embodied mind, where embodiment is compatible with the interactive unity of an independent mental substance and a body.

However, the ''too many thinkers'' objection destabilizes the Cartesian Composite View. The Cartesian mental substance is the primary source of thought. If we follow the rule for picking ourselves out, namely picking out the thing that is the non-derivative or primary source of the thought in us, then we will be the mental substances and not the composites. So, even if there are Cartesian composites, they are not human persons. It is the Cartesian mental substances, ''souls'' if you like, that will count as persons.

Perhaps then the best we can say is that in tracing the bodies associated with such souls, we were, by happy accident, tracing something strongly causally tied to the persisting persons. For all that can be set

out in a relatively a priori fashion, we might just be such independent mental substances. At least that is so, if they are lodged in animals more intimately and less adventitiously than, say, pilots in ships, to use Descartes' own example. For in watching a ship go by we are not tracking its pilot, unless we know that there is a pilot there. In tracing each other in the offhand ways in which we do, do we have such a working hypothesis, let alone knowledge, of mental substances driving the bodies before us?

The real difficulty with the idea that we include, or are, independent mental substances is the massive amount of evidence of the dependence of our mental functioning on the condition of our brains. The simpler empirical hypotheses that account for this all suggest that we are not *independent* mental substances. That leads to the next "mentalistic" version of what we essentially are.

Are we then associated with mental substances that emerge from the development of brains, and are dependent on the brains that they emerge from, and so suffer a loss in their capacity to function if those brains are damaged? Of all mentalistic views, perhaps the most natural is that animals come to be able to think by having some such mental substance develop in them. Such an account could be reconstrued as an account of what human animals consist of, besides their bodies. For, by hypothesis, the dependent mental substance ceases to exist with the destruction of the brain, and hence with the total destruction of the animal. Because the mental substance has no possibility of post-mortem existence and operation, we are not compelled to think of it as a thing which literally thinks, as opposed to a thing whose operations constitute the thinking of the animal. It is as if an animal is being taken to have a mental organ dependent on its brain, an organ whose operation constitutes the animal's thinking.

Such an account seems compatible with the view that we are human beings. (Even though it implies that my body is not wholly physical, it still provides adequate ontological underpinnings for the view that I am a human being.) For on this account, we would be contingently animals. We would go where our living brains, or heads, go. For those items are the loci of the physical complexity responsible for our mental substances and hence for the things whose operations constitute our thoughts. The brain is now conceived not as the chariot of a pure ego or Bare Locus, but as the essential physical underpinning of a dependent mental substance.

So the view that we are human beings is compatible with either materialism or an emergent mentalism that claims that any developed and undamaged human animal has a dependent mental substance as a part.

15

We now cross an ontological Rubicon, and consider a class of entities in a different *category*. Those already considered were substances. We will soon be dealing with cross-time bundles.

Now there are at least three philosophical conceptions of substance. The most demanding incorporates the idea that a substance must be ontologically independent, so that no level of ontological analysis will properly treat it as some other item modified a certain way. Set that aside. The least demanding characterization is that a substance is at the bottom of the hierarchy of predication; thus a substance is said to be ''a bearer of properties that is not itself a property''. That will seem insufficient to those who believe in events as well as substances, for events also seem to be at the bottom of the hierarchy of predication. I would say the same for species, understood as higher-order individuals; species are not predicated, it is membership in a species that is predicated. The same could be said for cross-time bundles.

The third conception of substance augments the least demanding characterization by requiring that anything that deserves the name of ''a substance'' has, at each moment of its existence, a power of self-maintenance, development, and persistence (at least relative to its natural environment) which would have to be cited in any adequate account of what it is to be the substance in question. Something like this seems to be behind Aristotle's idea that paradigm sublunary substances are 'the individual man and horse''. Consonant with the Aristotelian tradition, we may further require that the essence of a complex substance involves a form or principle of unity, a relation among its possibly varying parts which is such that if it holds at a given time then the substance exists at that time, and has a power of self-maintenance, development, and persistence (at least relative to its natural environment).[15] Given this, it

[15] For a more complete development of this view see ''Hylomorphism as a Theory of Unity'', forthcoming in *The Journal of Philosophy* (2006).

is also natural to require that anything that deserves the name of ''a substance'' persists by *enduring*, which is to say that it has all of its essence present at each time at which it is present.[16] In this respect substances contrast with *perduring* things, which are not such that all of their essence is present at each time at which they are present. This completes the conception of substance that is useful for present purposes.

In respect of their manner of persisting, substances are to be contrasted with what I will call ''cross-time bundles'' or ''successions'': items that persist from one time to another by having distinct phases of their total reality present at the two times. Think of a cross-time bundle in this way. It is present at any time at which one of its phases or temporal parts is present, in the same way in which I am present wherever one of my material parts is present. The bundle's being present at a time consists in its phase or temporal part existing at that time, just as my being present where my foot is consists in my foot existing there. Any cross-time bundle that is a candidate for being a human person essentially persists, it could not be an instantaneous existent, like an utterly punctuate event. Humans cannot have an instantaneous existence. So the candidate cross-time bundles consist of events and states that exist at different times, events and states tied together by a cross-time principle of unity. So if we restrict ourselves to the reality of such a bundle at a time, a phase of the bundle, we are not thereby capturing the essence of the bundle. What it is to be the bundle involves something more than what is going on at the given time. It lies

[16] This, and not the more familiar condition that an endurer be wholly present at each moment of its existence, is a better characterization of ''endurance'' as opposed to ''perdurance''; a now familiar terminology first introduced in my dissertation, *Particulars and Persistence* (Princeton, 1984). Unfortunately, my dissertation's gloss on endurance in terms of being wholly present, a gloss taken up by David Lewis in *Plurality of Worlds* (Oxford University Press, 1984), has now become standard. Given this account of endurance, it is unclear how a variably constituted entity, one that can gain and lose parts, can even be a candidate to be an enduring entity. (Are all of the parts it has over time to be present at each time? Of course not! But then which parts are to be wholly present at each time?) Yet many enduring substances *are* variably constituted.

The reader familiar with perdurance and endurance will see in the text that follows the germ of an argument to the effect that endurers are epistemologically more ready to hand, and probably epistemologically more basic than correspondingly *long-lived* perdurers. Short-lived perdurers, for example short events like noises and flashes, can simply capture our attention, without the subvention of a grasp of gen-identity conditions.

in the essence of such a cross-time bundle of phases that there should be more to it than what is present at an arbitrarily chosen time. The then present phase itself hardly constrains which bundles include it. That depends on what has happened elsewhere and else-when, and on the available cross-time unity conditions—sometimes called ''gen-identity relations''—that can bundle phases together.

So I would say that a cross-time bundle does not have all of its essence present at each time at which it is present, just as I do not have all my essence present at each place at which I am present. For example, not all of what is essential to me is present where my foot is. (In fact, none of what is essential to me is present where my foot is.) Not all that is essential to the history of New Jersey is present now. The cross-time bundle of states and events that is the history of New Jersey ''perdures'' through the times at which it has proper temporal parts, just as I perdure through those spatial regions where I have proper material parts.[17] Cross-time bundles perdure through time.

Having introduced the distinction between enduring and perduring, along with the claim that cross-time bundles perdure, we may now consider a bundle view that parallels the substance view of human beings described in the previous section. The phases are to be complexes of time-restricted physical states and events, states and events which obtain in the relevant person's body or brain at the time to which the phase corresponds. When the human being is enjoying some mental life, the complex phase associated with any such time will also involve the relevant mental states and events. So the required gen-identity relation will have to be able to unite the following types of phases into a cross-temporal whole.

[Pi,Pj, . . .], [Pl,Pm, . . . ,Mi,Mj, . . .], [Ps,Pt, . . .], [Pv,Pq, . . . ,Ms,Mt, . . .]

Where Pi, Pj, . . . etc. are (appropriate temporal restrictions of) physical states and events, and Mi, Mj, . . . etc. are (appropriate temporal restrictions of) mental states and events.

[17] So I think we should not say that I exist in the region occupied by my foot; and we should not say that the cross-time bundle exists at each time at which it is present. My foot exists there, and some or other phase exists at each time at which the bundle is present. But, for example, it seems wrong to say that this semester exists now; only its present phase exists now. Is this connected with our reluctance to say that momentous events are now existing, as opposed to now *occurring*?

Given that, the gen-identity requirement will have to be something like this: one phase [Xi . . .] is a phase of the same human person as another phase [Yi . . .] if and only if the physical constituents of [Xi . . .] are physically continuous with those of [Yi . . .], and the mental constituents of [Xi . . .] if any, are psychologically continuous with those of [Yi . . .], if any there be.

A human person can now be defined as a *maximal* bundle of such continuous phases. Every phase that stands in the above relation to any phase in the bundle is included in the bundle.

There are several things to note about this ontological picture of a human person. First, there is the complexity of the gen-identity relation, defined as it is in terms of physical continuity. The notion of physical continuity is to be further defined as the ancestral of tight connectedness, an immediate relation of dependence, which holds between successive phases in the bundle. Secondly, to know just what degree of connectedness to insist upon, we need to have already considered a wide range of cases in which a person would survive. For example, to properly demarcate the cross-time bundles that correspond to human beings, we would need to allow that there would be sufficient physical connectedness among a person's phases when he is, all of a sudden, surgically reduced to the condition of a head.

Knowledge of such relatively sophisticated empirical matters is not plausibly taken to be part of what we are relying upon when we trace each other through time. Instead, we appear to offload such issues onto the persons themselves. In tracing persons we are deploying knowledge of *some* necessary conditions for their survival over time—they can't explode into smithereens, for example—but we need not be relying on knowledge of any non-trivial sufficient condition for their survival over time. In order to trace persons, we need not know what condition suffices for their persistence in every future or possible situation.

We can offload onto substances, for at any time at which they capture our attention, all of their essence is present before us; and so at any such time the basis of their natural power to persist is present. Without taking any specific view of what that basis consists in, or even any specific view of the conditions under which it would persist, we can rely on there being such a basis. So even though we do not bring to bear any sufficient condition for the identity of a substance over time; when a substance does capture our attention, there is something in that substance at that time, namely its natural power to maintain itself in a

certain way, which settles the question of *what it would be* for the thing which has captured our attention to persist. As the earlier example of the ''android'' who was really human brought out, that question is settled by the nature of the substance; it does *not* depend on what cross-time identity conditions *we* are deploying to trace the substance.

Can we offload in this way onto cross-time bundles? Suppose that a phase of such a bundle now captures our attention. There is nothing in that phase which itself selects one cross-time bundle from a host of others as the bundle to which it belongs. There is an enormous variety of gen-identity relations which could connect that bundle to this, that, or the other sort of phase in the future. There are an enormous variety of past and present things whose phases could be bundled together with any given phase that is present to us at a time. Even with respect to a specified series of evolving phases, more restrictive gen-identity relations determine shorter-lived bundles, and less restrictive gen-identity relations determine longer-lived bundles. The present phase itself determines none of these bundles. So which bundle are we tracing? Absent our seeing the present phase as a phase of a bundle of a definite kind, there is no definite answer to that question. And if we are to see the present phase as a phase of a bundle of a certain kind then we need to have some gen-identity condition in mind.

On the bundle view of persons, the object of our attention when we glimpse a person is a phase, and by association some bundle which contains that phase as its present part. But the phase radically underdetermines the content of, and hence the answer to, the question: What would it be for the bundle that is the object of attention to persist or fail to persist through time? There is no such thing as *the* bundle we are attending to; there are as many such bundles as there are gen-identity conditions that apply to the phase.

If in tracing objects and persons through time, we were implicitly or explicitly deploying specific gen-identity conditions, then this would determine a content for, and an answer to, the question: What would it be for the bundle to which I am now attending to persist or fail to persist?

The claim that we are implicitly or explicitly deploying specific gen-identity conditions in tracing objects and persons through time is just the denial of the plausible hypothesis of offloading.

This implies that if human beings were the cross-time bundles introduced above, then we would not have been tracing them, when in our

easy and offhand ways, we trace ourselves and each other. And that would suggest that we are not human beings, after all.

That conclusion can be avoided, thanks to the availability of a plausible conception of human beings as substances. A human being is a substance that begins life either as an embryo or a fetus; and as such, it has the natural power to develop into a grown animal whose brain becomes the material basis of a person's mental life. All of that potentiality, quasi-miraculously, is encoded in the embryo or the fetus at any time at which it exists. The animal that develops in the course of the natural development of the fetus has at each time at which it exists the natural power to maintain itself as, and so persist as, an animal. (Its essence involves the power to sustain itself as such in its natural environment.) The person that develops in the course of the animal's development also has the natural power to persist as such, until the animal's brain is severely degraded. And this natural power to persist would also operate in those artificial situations in which the animal's physiology can be medically replaced or augmented. So on this conception a human being as a substance, at any time at which it exists, has the capacity to survive as a separated head, artificially kept alive and functioning.

You diehard Carnapian conventionalists out there take note! Even though there is a mapping from a substance conception of human beings to a bundle conception, the substances can play an epistemological role that the bundles cannot.

The question arises whether the items deemed persons by the Wide Psychological View can be understood as substances, or only as cross-time bundles.

16

The Wide Psychological View first came clearly into focus because of an observation of Jerome Shaffer's, which was followed up by David Lewis.[18] The observation was that a Materialist, a theorist who regards persons as wholly constituted by their (wholly material) bodies, could

[18] J. Shaffer, ''Persons and Their Bodies'', *The Philosophical Review* 75 (1966) and D. K. Lewis, ''Survival and Identity'', in A. O. Rorty, (ed.), *The Identities of Persons* (Berkeley: California UP, 1976).

nonetheless consistently allow that persons can survive the destruction of their bodies! (The resurrection of the dead is thus possible without the resurrection of the actual bodies of the dead; so all that theological subtlety about what happens when at the general resurrection when God has to resurrect even a third-generation cannibal whose body was made up of flesh made entirely from the bodily elements of others, goes by the board.[19])

We can, after all, distinguish *synchronic* and *diachronic* unity conditions for a person. A Materialist does need a purely bodily *synchronic* unity condition for (human) persons; he needs to maintain that at any time a person is wholly constituted by his body. But this seems to leave the conceptual possibility that over time different bodies could constitute the same person. Well, what then would make it the same person who had those different bodies? The answer was given by Shaffer and Lewis. The different bodies could be such that their (physically reducible) mental lives were psychologically continuous.

Isn't this precisely what is depicted in the fantasies of teletransportation, as in *Star Trek* and the like? Maybe not; maybe the teletransporter is just an independent Cartesian mental substance that moves from one body to a new body made in the old body's likeness. Set that aside, since our total evidence concerning the dependence of mind on the brain counts against such independent mental substances. Still, the Wide Psychological View can easily accommodate the idea that we would survive teletransportation; for the body-copying that it incorporates does secure psychological continuity. The person in the newly created body has memories of the acts and experiences of the person who stepped into the teletransporter, and he intends to carry out the plans of that person; his beliefs, outlook, and character rather exactly correspond as well. On the Wide Psychological View, this is sufficient for him to be the same person as the one who stepped into the teletransporter.

The Wide Psychological View thus implies that we are the sorts of things that can survive bodily destruction and replacement. It also implies that we are the sorts of things that survive a brief interlude of *not existing*. After all, the passing of the signal from the teletransportation booth to the supposed new destination of the person, takes a finite

[19] The ''solution'' was that at the general resurrection of the dead, we all will come back slightly shorter than we were so as to accommodate the shortage of parcels of flesh due to all the cannibals there have ever been.

amount of time. During that time, the person who is supposed to survive the overall process has no bodily or psychological reality! Nor is he constituted by a signal, for the signal is not the sort of thing that could be psychologically (or physically) continuous with a human person. The supposed survivor does not exist during that period.

So here is a not-widely-noticed consequence of the Wide Psychological View: we can intermittently exist. This in itself is not an objection to the view. However, it is worthwhile to investigate the sorts of things that can be intermittent, and ask how it is that they can be so.

Performances can be intermittent. During the interlude between the first and second acts of a performance of *Tristan und Isolde*, the performance is not taking place. It has (mercifully, some might say) ceased for twenty minutes or so; it will begin again with the prelude to the second act. In the last raucous and disorienting moments of the performance of the first act, there is not something with all of its essence present which *is* the performance, and which incorporates the natural power to generate the prelude to the second act. That power lies in the conductor and the orchestra, who will play the music, and in the stage hands and the stage manager, who will see that the curtain opens upon "a garden with tall trees in front of Isolde's apartment". The performance can be intermittent because these continuously existing items generate it.

So, the performance is not a substance. It does not contain within itself, at each moment while it is going on, a power of self-maintenance, development, and persistence which would have to be cited in any adequate account of what it is to be that performance.

The performance should instead be modeled as a cross-time bundle of phases made up by events and states, phases united by a gen-identity condition that might say something like: one phase is a phase of the same performance of *Tristan und Isolde* as a later phase if and only if they are in the ancestral of the relation of being the immediate next phase of a performance of *Tristan und Isolde*; where one phase is the immediate next phase of a performance of *Tristan und Isolde* if and only if the same company (of singers, musicians, and stage hands) produced both phases, and the later phase followed in sequence from the earlier in accord with the score, libretto, and stage instructions of the work *Tristan und Isolde*.

So there will be a non-trivial sufficient condition for the last phases of Act One of a particular performance to be stages of the same performance of *Tristan und Isolde* as the first phases of the prelude to Act Two. And that condition will not require that the last phase of Act One be

anything like a substance, whose always present essence would involve a power of self-maintenance, development, and persistence that will itself under normal conditions bring about the first phases of Act Two.

We can distinguish two ways, among many, of attending to *what is going on* during a performance of *Tristan und Isolde*. One may simply be attending to the music, the singing and the acting; simply drinking it in, however it happens to go. (A child may do this at her first opera.) Or one may also be tracking this particular performance, and evaluating *it*, say by comparison with near ideal performances, such as the fabled June 1937 Covent Garden performances with Lauritz Melchior and Kirsten Flagstad, conducted by Sir Thomas Beecham.

Notice that the second sort of audience member, but not the first, would have *made a mistake* if he had sat down at the Met expecting the performance of an opera, and the conductor had raced the orchestra through the prelude to the first act of *Tristan*, but only as the introduction to a lecture about the music of Wagner. For this auditor would have heard the prelude as the opening of a performance of the opera, and perhaps doubted whether the same frenetic tempo could be kept up throughout the whole performance.

If one's object of attention includes not just the music, the singing, and the acting, but the unfolding performance of the opera, then one has to be seeing and hearing the phases of the opera *as* parts of a performance of an opera. And there is no way to do that without deploying knowledge of what a performance of an opera is. And that involves at least implicit knowledge of what is sufficient for one phase of such a performance to be a phase of the same performance as another phase.

I believe that what here applies to performances applies to cross-time bundles of phases in general. To trace such a bundle is to perceive phases in the bundle *as* parts of the same bundle. And that involves at least implicit knowledge of what is sufficient for one phase of such a bundle to be a phase of the same bundle as another phase.

I also believe that the best explanation of what we are, consistent with the surprising (and I think false) claim that we can intermittently exist, is that we are psychologically continuous cross-time bundles. A substance, by contrast, cannot come back from the existential grave. Even if one substance causes an exactly similar substance to come into existence a few seconds after the first has ceased to be, that is one substance being replaced by another.

I also believe that in tracing ourselves and others through time, we are offloading; that is, we are relying on the power of self-maintenance,

development, and persistence of salient substances, rather than deploying sufficient conditions for bundling phases together. Certainly we trace people even when we are at a very early developmental stage, say by the age of two, just to opt for a quite conservative estimate. We do not then have knowledge of such sufficient conditions.

Anyone with these beliefs should reject the Wide Psychological View. To make the argument explicit:

16. In tracing ourselves and other people through time, we are frequently offloading.
17. Given its toleration of intermittent existence, the Wide Psychological View should be construed as the view that we are certain sorts of psychologically continuous cross-time bundles.
18. We cannot trace such a bundle by offloading.
19. So, the Wide Psychological View gives the wrong account of what we are tracing when we trace ourselves and others.
20. So, the Wide Psychological View gives the wrong account of personal identity.

We need not make things rest on the consideration of intermittent existence. Here is an empirical fact: there are no independent mental substances. So any substance associated with a person, be it a body, an animal, or an emergent but dependent mental substance, would cease to exist if the whole material reality of the person were destroyed. But the whole material reality of a person who gets into a teleporter is destroyed. So if the Wide Psychological View is to preserve its distinctive claim that we can survive teletransportation and the like, then it must view persons as cross-time bundles. Yet, in tracing persons we offload onto the persons themselves. So we, the persons that we have been tracing and re-cognizing, are not cross-time bundles. The Wide Psychological View does not apply to us. But *we* are the persons who are the subject matter for the philosophy of personal identity. So the Wide Psychological View should be rejected as an account of personal identity.

17

Just as the Animalists are wrong to identify me with something essentially constituted by the H. M. S. Johnston, the friends of the Wide Psychological View are wrong to take me to be a maximally

temporally extended Johnston-*esque* performance or bundle, or anything of that sort.

I am a human being. Well, more precisely, at a certain level of ontological analysis appropriate to the kinds of considerations that divide the Animalists and the friends of the Wide Psychological View, the thing to say is that I am a human being.

Even so, in the light of other considerations not appropriately canvassed here, it might emerge that our natures are more Protean than we think, so that ''human being'' turns out to be a crypto-phase sortal, just like ''human animal''.[20]

[20] For a defense of this position, which does not depend on endorsing the Wide Psychological View, see my ''Relativism and the Self'', in M. Krausz (ed.), *Relativism: Interpretation and Confrontation* (University of Notre Dame Press, 1990).

Part II

MODALITY

This page intentionally left blank

3. Worlds, Pluriverses, and Minds

Mark Heller

Over the last few years I have been developing an ontology of ersatz possible worlds based on some suggestions made by W. V. O. Quine.[1] The proposed account identifies worlds with complicated sets that represent distributions of fundamental properties across a manifold.[2] I will henceforth call this view "Representationalism". The purpose of the present chapter is twofold. First, I want to consider how this picture of worlds must be developed in order to accommodate the possibility of manifolds that are not connected with one another or only partially connected with one another. Secondly, I want to consider how this picture of worlds must be developed in order to accommodate the possibility of non-physical minds. The relation between these two projects is that I will propose that non-physical minds can be treated as collections of mental properties distributed in separate manifolds each of which is partially connected to the manifold within which the physical properties are distributed.

I begin in Section I with an account of the Representationalist theory of worlds. In Section II, I extend the theory to include disconnected and

[1] W. V. O. Quine, "Propositional Objects", in *Ontological Relativity and Other Essays* (New York: Columbia University Press, 1969), 139–60. Quine proposes an account of worlds according to which a world tells us which points in space are empty or full at any given time. He does not actually accept this account of worlds, since it attributes to every world features that science seems to tell us our world doesn't even have (p.152).

[2] See my "Property Counterparts in Ersatz Worlds", *The Journal of Philosophy* 1998, 293–316, and "Five Layers of Interpretation for Possible Worlds", *Philosophical Studies* (1998), 205–14. The kernel of the ontology I develop was explicitly considered and rejected by David Lewis in *On The Plurality of Worlds* (Oxford: Basil Blackwell, 1986). I answer those objections in "Property Counterparts in Ersatz Worlds". Ontologically worlds are just sets, but they do represent distributions of fundamental properties, so in addition to the existence of sets, I am committed to the actual existence of that number of properties that the actual world represents the concrete world as containing. The ontology of worlds that I propose will be found controversial in light of its being part of an overall conventionalist program with respect to ordinary objects. To the extent that my development of the ontology of worlds within the conventionalist framework is deemed successful, some support is provided for that framework.

partially connected manifolds. In addition, this section will serve to clarify a challenge to David Lewis's modal realism.[3] Section III will explore the way minds can be fit into this picture of worlds.

I. WORLDS

There is a well known ambiguity in the word "world", especially noticeable in the expression "actual world".[4] Representational worlds are abstract objects that represent the one concrete world. The concrete world is not a possible world at all, and in particular it is not the actual world. The concrete world is composed of me and you and stars and atoms. We can bang on pieces of it, but we cannot bang on "pieces" of the abstract worlds that represent it. The accurate representation is the actual world. All the other abstract worlds are merely possible. They misrepresent the concrete world. I will henceforth reserve the word "world" for the abstract representations except when I explicitly modify it with "concrete" in order to refer to the thing we can bang on.

A world is a representation of a distribution of fundamental properties across a manifold. We represent the fundamental properties with numbers.[5] We represent the manifold with ordered n-tuples of numbers, which serve as coordinates picking out a specific location in the manifold. We represent one or more fundamental properties as being instantiated at a location with an ordered pair, the first member of which is the location representer and the second member of which is a set containing all of the relevant property representers. A world is a set of these instantiation representations.[6] A very simple world with a discrete manifold might be visualized like this:

[3] Lewis presents his view in *On The Plurality of Worlds*. All page references in the body of the chapter are to that text.

[4] See, for instance, Robert Stalnaker, "Possible Worlds", *Nous* (1976), 65–75.

[5] In the body of the chapter I use numbers for convenience. But we should remember that numbers themselves can be identified with, or at least replaced by, sets constructed from the null set. So it will turn out that my worlds are just sets of sets constructed from the null set.

[6] To be more careful, a world is an equivalence class of these sets to allow for the arbitrariness of any given coordinate system. Note that the coordinate systems do need to be equivalent with respect to the betweenness relations among the locations. Presenting the view in terms of coordinate systems makes it easier to grasp, but it does unacceptably limit the sorts of manifolds that can be represented. A more acceptable, though less envisionable, option would be to let the manifolds be represented by mathematical

{<<0,0>,{1,2}>,<<0,1>,{1}>,<<1,0>,{2}>}.

This world represents two fundamental properties distributed across a two-dimensional manifold containing just three locations, each of the properties being instantiated at two of the locations and being co-instantiated at just the one location at which the two dimensions intersect.

The world I have presented represents two properties distributed across a manifold, but it does not specify which properties those are. The numbers do not, in themselves, represent any particular properties. What makes a number represent the property of, say, negative charge, in a given context is our conventions for identifying negative charge. We look at the role that a property plays in a given world, and if that role counts as similar enough, in that context, to the actual role of negative charge, then the number in question represents negative charge in that context. There are different features of a property's role within a world that might receive more emphasis in different contexts. For instance, we may focus more on the property's distribution and how that compares to the distribution of negative charge in the actual world. Or we may instead focus more on how that property fits into the laws of its world and how that fit compares to negative charge's fit into the laws of our world. Or we may try to balance these two factors.[7]

The representation relation, in general, is context sensitive, and it is no different for the case of the numbers representing properties. Nothing surprising here. But there is a way to make it sound surprising. Ask not about the representers but instead consider the properties being represented. If a number in another world represents the same property

formulae that describe the structure of the manifold. Then we can pair those mathematical representations with functions from locations in the described manifold to numbers representing properties at those locations. This has the added advantage of segregating one area of potential metaphysical dispute. Though I make assumptions about the possible natures of manifolds throughout this chapter (for instance, that the temporal dimension is not intrinsically temporal, but only temporal in virtue of the way properties behave in that dimension), I have no great stake in such assumptions. I am prepared to leave those debates about the manifold, and, therefore, debates about what will count as an adequate mathematical representation of a manifold, to those who are better equipped to handle them. My concerns are with the metaphysics of modality and the metaphysics of objects. I thank Jim Pryor for pointing out to me the advantages of this alternative method of representation.

[7] For more on the balance between similarity of distribution and similarity of nomological fit see "Property Counterparts in Ersatz Worlds".

that exists in this world, negative charge for example, then this looks like a straightforward case of transworld identity of properties. The property that that other world is representing as existing in that world is the very same property that exists in the actual world. But this representation only occurs at a very high level of interpretation of the sets doing the representing.[8] At a lower level of interpretation, the numbers represent properties without representing them as any specific properties. The properties being represented at this lower level of interpretation are not identical to any properties in any other world. The world represents a property playing such and such a role, but until that world is further interpreted, the property is not represented as being negative charge or any other property. The role that property plays in its world is more or less similar in various respects to the role played by negative charge in the actual world. It is in virtue of these similarity relations that we end up (at a higher level of interpretation) treating, or not treating, the property of the one world as being identical to negative charge. Moreover, which respects of similarity are deemed relevant and whether there is enough similarity in the relevant respects to warrant treating that otherworldly property as negative charge are context sensitive matters. This emphasis on similarity and context makes the transworld relation between properties look more like the counterpart relation than the identity relation. And that does sound surprising.[9]

Objects also can be understood on the counterpart model or the identity model, depending on which level of interpretation we are considering. When thinking of a world as numbers representing unspecified properties, it is only by interpretation in accordance with our conventions that regions at a given world end up counting as containing objects at all, and it is only in virtue of our contextually sensitive emphases on various respects of similarity that an object at another world gets to be treated as being the same object as one at the

[8] For more on levels of interpretation see ‘‘Five Layers of Interpretation’’.

[9] There is this much reason for thinking it is better to think of the transworld relation as the counterpart relation rather than identity: the relation is grounded in similarity which is neither transitive nor one–one, and identity must be transitive and one–one. But it could instead be argued that in any context in which an identity is represented the only possible worlds involved in that representation of transworld identity will be ones across which the relation in question will be transitive and one–one. See my ‘‘Interpreting Worlds: Identity and Impossibility’’, *Philosophical Topics* (2002), 77–101.

actual world. Yet, when we consider the world as fully interpreted, it may well represent actual world objects as having properties they do not actually have. At this level of interpretation, it is the very same object being represented as existing in the other world as the one that actually exists in our world. At this level of interpretation there is transworld identity.[10]

One important respect of transworld similarity for objects and properties is, in many contexts, how those objects and properties fit into the laws of nature. In the spirit of complete modal reduction we can be Humeans about laws. In this case the laws of a world are themselves a function of the property distribution of that world, together with our principles of simplicity and elegance and the like. Just as the laws of the actual world can be a function of the actual world's property distribution independent of which otherworldly properties the actual world's properties are identical to, so other worldly laws can be a function of that world's property distribution independent of transworld identity issues.

[10] If we want to focus on worlds at the highest level of interpretation, we can construct fully interpreted worlds out of the basic worlds I have provided. Full interpretation is relative to a language and a context. A single basic world may contain cars when described in English but only molecules in that same location when described in a language that lacks words for macroscopic objects. These are just two descriptions for the same facts. (Here's that conventionalism.) In *The Concept of Identity* (New York: Oxford University Press, 1982), Eli Hirsch has presented a language that has the words "incar" and "outcar" instead of "car". An incar only exists inside a garage, and an outcar only exists outside a garage. As an incar approaches the garage exit it shrinks, and an outcar begins to grow just across the threshold. This language, too, can be used to describe those same facts. These three languages yield three different fully interpreted worlds from the single basic world.

We can construct each of these fully interpreted worlds by joining in an ordered pair the basic world with all of the sentences of the language that, when combined, describe the same facts as those described by the basic world. But this is not quite good enough. We also need to take account of the role of context. A single basic world might be interpreted relative to one context as containing Heller but as containing a mere Heller look-alike when described relative to another context. The same basic facts—the same distribution of fundamental properties—are subject to either of these interpretations. To get fully interpreted worlds we pair the basic worlds with all of the sentences of a given language that in a given context describe the same facts that the basic world describes. So, even if we restrict our attention to English translations of a single basic world, there will be many fully interpreted worlds for the one basic world. Shifts in context do not affect which of these worlds exist, but such shifts do affect which are possible. Some fully interpreted worlds contain Heller as a talking car. Whether these worlds are possible or not depends on what it takes to count as Heller in a given context. This sort of context relativity is just what we should expect from an anti-essentialist like myself. It's just what we should expect when we ground transworld identity in similarity plus an interpretation. For more on transworld identity and the context relativity of impossibility, see "Interpreting Worlds: Identity and Impossibility".

We are not stuck in a circle in which a world's laws depend on the transworld identities of its properties and the transworld identities depend on the world's laws.

I confess a certain attraction to a less than complete modal reductionism. Though I have no argument for it here, I am prepared to consider brute laws of nature, though I fully disclaim brute essential properties. I find attributing brute modal properties to a thing to be more repugnant than brute laws. If I want to add brute laws to my worlds I can do so, representing them with functions from property distributions to property distributions.[11] Two worlds may be identical with respect to the property distributions at those worlds but differ with respect to the functions. Such a difference will in turn ground different counterfactuals at the two worlds. I do not here endorse this view. I mention it only in case I wish to endorse it at some later date.

So worlds are sets. Not just any set is a world, but only those sets that have the structure that I have described. They must be able to represent properties distributed across manifolds. Are there sets with the right structure that are not worlds? I say "no", and I appeal to the principle of recombination. The principle of recombination is the principle that gets stated in pre-philosophical discourse as "anything is possible". If the question were which objects can be combined with which, the answer would be any and all (provided the combination is consistent, by which I mean at least: provided the combination does not violate our conventions for identifying the objects in question). But that is not the question, because objects are just conventional constructs out of properties. If the question were which properties could be combined with which, the answer would be any and all (provided the combination is consistent, by which I mean at least: provided the combination does not violate our conventions for identifying the properties in question). But that is not the question either, because transworld property identity is itself a conventional construction out of patterns of property distribution. The correct question is which property distributions are possible, and the answer is any and all. Put in terms of the sets that are doing the representing, the principle of recombination tells us that any set that has the right structure is a possible world. If we can represent a distribution

[11] The functions themselves can be represented by sets of ordered pairs of property distributions, the first representing the argument of the function and the second the value of the function.

of properties, that distribution is possible. Furthermore, if it should turn out that stating the principle of recombination in terms of combinations of objects instead of distributions of properties leaves out some distributions, then for that reason the statement in terms of objects is inadequate.

So every set with the right structure is a world. But there is still the worry that the structure I have selected is too restrictive, the worry that there are possibilities that I cannot represent. I am not worried about the possibility of talking donkeys. Even though I do not represent donkeys explicitly, talking or otherwise, they do get represented implicitly by representing a distribution of fundamental properties which, by applying our conventions to that distribution, is a representation of a talking donkey. I do not have to be able to represent donkeyhood in addition to the fundamental properties. But I am worried about the possibility of minds. If there were non-physical minds, they would not be located in space, the way donkeys are. But, so far, I only have the ability to represent things at locations, and if donkeys and minds existed they would have to be represented as being some spatial distance from one another, which would require the minds to be in space. That problem also suggests a separate but related worry—the possibility of spatial entities that are not spatiotemporally related to each other. These objects would be in space, but not in the same space. So, independent of any question about minds, I need to find a way to represent these unusual manifolds. It is to that challenge that I now turn.

II. PLURIVERSES

I am not alone in being threatened by the possibility of unusual manifolds. Lewis's modal realism has been attacked on similar grounds. The objection has been made in different ways. One way to say it is that even if there are all the non-connected concrete universes Lewis says there are (what he calls "worlds"), all of those universes are part of a single world, a single way that things can be. Put another way, even if there actually are all these universes, there might not have been. The world might have contained a pluriverse composed of fewer universes, or it might have contained just one universe. For the time being, let us set aside the question of whether this is an overwhelming, or even fair, objection to modal realism, turning instead to the worries that this style

of objection raises for my theory of worlds as it has so far been presented.

My account of worlds so far does not allow for the possibility of many universes within a single pluriverse. This is because my worlds are limited to a single manifold. Thus, everything within a single world will exist within the same manifold. To accommodate the possibility of pluriverses, I must allow for disconnected manifolds. So I revise my theory by adding superscripts to every set that represents a location.[12] Each superscript represents a different manifold. So now I can represent disconnected manifolds within a single world, and I can represent any number of disconnected manifolds; it is not all or none. For instance,

$$\{<<0,0>^1,\{1,2\}>,<<0,1>^1,\{1\}>,<<1,0>^1,\{2\}>,<<0,0>^2,\{1,3\}>,<<0,1>^2,\{3\}>\}$$

represents a pluriverse containing two universes, the first of which contains three locations and two properties and the second of which contains two locations and two properties, one of which is the same as a property contained in the first universe.

And there are more possibilities that I can now represent that I had left out before. Suppose that instead of superscripting every location-representing set, I instead superscript every number within the location-representing set. Now I can represent universes that are connected along one dimension but not along others. (I don't care if we call these two partially connected universes or one partially disconnected universe. I will tend to use the former terminology, but not for any reason.) For example we might consider two universes containing spacetime manifolds. If the laws governing the pluriverse are such that the dimension shared by the two universes plays the role of the temporal dimension in both universes, then we can describe the pluriverse by saying that the two universes are connected along the temporal dimension but not the others. But it could turn out that the laws are such that the shared dimension is the temporal dimension in one universe and a spatial dimension in the other. That is one more possibility that a complete theory of possible worlds should be able to capture.

$$\{<\{0,^1,0^2\},\{1,2\}>,<\{0^1,1^2\},\{1\}>,<\{1^1,0^2\},\{2\}>,<\{0^3,0^2\},\{1,3\}>,<0^3,1^2\},\{3\}>\}$$

[12] Of course, adding superscripts is just a typographical shorthand for constructing ordered pairs.

represents a pluriverse containing two two-dimensional universes that share one dimension (or, if you prefer, it represents a single three-dimensional universe in which two of the dimensions do not intersect).[13]

Perhaps wholly disconnected universes within a pluriverse can causally interact. (I say "perhaps" because I think that the considerations of this and the next paragraph are, though suggestive, not wholly convincing.) Perhaps two universes with their own spacetimes, for instance, can still be such that what obtains in one of the universes is causally relevant to what obtains in the other, though only if we understand causal terminology as not analytically implying temporal relations. First imagine a single, wholly interconnected universe with the following sort of laws: there are some conditions that obtain at all times and all places throughout the universe, but they only obtain because of some particular condition that obtains at at least one time and one place. It does not matter whether we think of this law as brute or Humean. It might be somewhat harder to imagine how there could be such a law on the Humean picture—harder to imagine what the rest of the universe would have to be like to make the best systematization of the particular facts be one that posits such a law—but I doubt that Humeans want to rule out such laws on principle.

So, let us suppose such a universe, with a property distributed throughout all of spacetime, call that property "orange", and another property, call it "yellow", that is instantiated at just location L, and suppose yellow to be a nomologically necessary condition for orange. Yellow's presence at L is causally relevant to orange's presence throughout the universe, including locations that are both distant from L and prior to L. If such a universe is possible, and if pluriverses composed of disconnected universes are possible, then so too should be pluriverses in which yellow's presence at L in one universe is causally relevant to orange's presence throughout the entire pluriverse, including locations that contain no dimensions that intersect L. The thought here is that if yellow's mere existence is what is relevant to orange, regardless of where it exists along any of the dimensions that might intersect an

[13] The location-representers no longer need to be ordered, since the superscripts keep track of which dimension is which. In truth, we don't need to order the location-representers with respect to the property-representers either, since it is clear from the structure which ones represent which.

orange place, then that existence should be, or at least could be, just as relevant even if it is at a point that doesn't intersect any of the dimensions in which some of the orange exists.

The possibility of causal interaction between universes that are partially connected is even more plausible. Imagine two spacetimes that share a common time though not any common space. Suppose that when and only when an event of type E occurs in either universe an event of type E^* occurs at two random spatial locations, one in each universe. Even a Humean should have no trouble imagining the rest of the world being such that the laws connect E events in either universe with E^* events in both universes in such a way as to make it true that the two universes are causally connected. The case may be even more convincing still if we suppose that the ''effects'' are somehow spatially correlated with one another (without bearing spatial relations to one another).

Let us flesh out the case. Suppose that the two universes are such that properties P_1–P_6 have the exact same pattern of distribution throughout the two universes. With respect to those six properties the universes are indistinguishable. For convenience, let's call these six properties the ''color properties''. The matching distributions of color properties, given a sufficient number of them at least, provides a sense in which the spaces of the two universes are correlated. With a complex enough pattern of colors, we can describe positions within that pattern by appealing to just the colors and not the locations at which the colors are instantiated. Since the two universes have duplicate color patterns, we can describe ''duplicate'' positions within those patterns. But now let us also stipulate that properties P_7–P_9, call them the ''waves'', are instantiated in different patterns throughout the two universes. (Put imagistically, it is as if the universes are two identical maps—identical by virtue of the colors—with the waves positioned differently on those maps.) Now let us suppose that occurrences of various combinations of the waves in one universe are followed by occurrences of various combinations of the colors in both universes (for instance, whenever P_7 and P_8 are co-instantiated in either universe, P_3 gets instantiated in both universes). Let us further suppose that the color or colors that become instantiated do so not at a random location, but at the same location at which the earlier wave properties were instantiated and at the correlated location in the other universe. I think that even a Humean would be willing to

attribute a causal relation between the waves and colors even across universes.[14]

I have described several different kinds of pluriverses, not just the simple case of two or more universes that have wholly separate manifolds. By superscripting the location-representing numbers in the sets which are my possible worlds, I can accommodate all of these various possibilities. The fact that there seem to be so many variations on the pluriverse theme adds more support to the original hypothesis that the simple pluriverse story really does describe a possibility. More persuasive support for the possibility of pluriverses, at least to my mind, is that we now have a way to represent such a possibility. The principle of recombination told us, in effect, that anything that could be consistently represented was a genuine possibility. It is not just that there are no limits on the possible combinations of objects. It is also that there are no limits on the possible structures of manifolds. Once that is taken seriously, pluriverses are easy to get, and any acceptable account of possible worlds must be able to count them as possible.

And now we can tighten up the challenge to Lewis. There are possibilities a representationalist ersatzist can account for that a Lewisian modal realist cannot.[15] Here are two versions.

[14] My trick for matching cause with effect in distinct universes is to find a correspondence relation that doesn't depend on spatial relations between the cause and effect. The two partially connected universes in my example are similar enough that we can impose a mapping from the one to the other. A problem with this strategy is that the correspondence relation I've found need not be one–one. Expand my example so that the pluriverse contains the two universes I have already described, call them "A" and "B", plus two more universes that are duplicates of the first two, call the two added universes "A^*" and "B^*". The event in A that supposedly causes two events, one in A and one in B, is mirrored by an event in A^* that causes an event in A^* and an event in B^*, each of which duplicate the effects in A and B. The problem now is that there is nothing that makes it the case that the cause-event in A is causing the effect-event in B rather than the effect-event in B^*. I thank Jaegwon Kim for pointing out this difficulty and Robert Howell for raising a similar difficulty. I note in response that the same difficulties will arise even within a single universe unless we require either something like spatial contiguity of the cause and effect (as considered by Kim in "Causation, Nomic Subsumption, and the Concept of Event", *Journal of Philosohy* (1973)) or require that something literally be passed from cause to effect (in keeping with Douglas Ehring's proposal in *Causation and Persistence* (New York: Oxford University Press, 1997)—though this solution ends up needing to posit brute diachronic identity for the item that gets passed).

[15] Phillip Bricker proposes a version of modal realism that allows for wholly disconnected pluriverses, but only at the expense of taking actuality to be a brute property of some concrete worlds but not others. See his "Island Universes and the Analysis of Modality", in G. Preyer and F. Siebelt (eds.), *Reality and Humean Supervenience: Essays on the Philosophy of David Lewis* (Rowman and Littlefield, 2001).

Version one goes like this. There are possibilities of partially disconnected universes within a single pluriverse. If there are those possibilities, then a sufficiently general principle of recombination, one that allows for the restructuring of manifolds, tells us that there should also be the possibility of wholly disconnected universes within a single pluriverse. Lewisian modal realism cannot accommodate this possibility, since such disconnected universes would count as separate possible worlds within the modal realist's framework.

Lewis might respond by questioning even the possibility of partially connected universes. He might raise doubts about my use of imagination in defending the possibility of partially disconnected universes. This would be an extension of such a worry that he raises with respect to imagining wholly disconnected universes:

> If you thought, as I did too, that a single world might consist of many more or less isolated world-like parts, how sure can you be that you really had in mind the supposed possibility that I reject? Are you sure that it was an essential part of your thought that the world-like parts were in no way spatiotemporally related? Or might you not have had in mind, rather, one of these substitutes I offer [earlier on p. 72]? Or might your thought have been sufficiently lacking in specificity that the substitutes would do it justice? (*On the Plurality of Worlds*, 72–3.)

I am not sure that Lewis would extend this objection to partially connected worlds. He does think both that for every Representationalist world as Quine originally described them there should be a corresponding Lewisean world (p. 90) and that Quine's original account of such worlds "could be extended to allow for various sizes and shapes of spacetime, for occupancy by different kinds of matter and by point-sized bits of fields, and perhaps even for occupancy of times by non-spatial things" (p. 90). I see my account of worlds as just the sort of extension Lewis would have had in mind here.

So, perhaps Lewis would instead object that the principle of recombination does not take us from partially connected universes within a single world to wholly disconnected universes within a single world. But I can find no plausibility to this answer. Once we accept the sorts of extensions to Representationalism that would yield the possibility of partially connected universes, there are more elements on the table to be subject to potential recombination. It would be wholly ad hoc to limit the principle to the recombination of objects and to avoid the recombination of the structures of manifolds.

I return then to the claim that mine is a questionable use of imagination. I deny this. I claim rather that my use of imagination is precisely the use that Lewis endorses when he asserts: "We get enough of a link between imagination and possibility, but not too much, if we regard imaginative experiments as a way of reasoning informally from the principle of recombination" (p. 90). I think that that is exactly what I have been doing with my imaginings. But whereas Lewis is limiting the recombinations we are allowed to imagine to recombinations of objects, I also allow recombinations of the structures of manifolds. This brings us to the second version of the challenge to modal realism; this version appeals directly to the principle of recombination, and leaves out any attempt at imagining alternative possibilities.

Version two goes like this. The principle of recombination itself, once broadened to account for recombinations of more than objects, tells us that there should be the possibility of wholly disconnected universes. The fundamental principle is that nothing consistent should be ruled out as a genuine possibility. But even that requirement would not rule in all the possibilities if the consistency were limited to a weak language that could not represent all the possibilities. What we want is a language with complete representational power. Then, and only then, the fact that nothing consistent is ruled out will be sufficient for ruling in all possibilities. A principle of recombination stated in terms of objects alone leaves out consistent recombinations that do not involve objects. Once we expand our representational ability to include differently structured manifolds, we allow a more general, a more thoroughgoing, a less limited, principle of recombination. And that more general principle is what demands the possibility of disconnected universes—a possibility that Lewis cannot allow us.[16]

III. MINDS

I now have the apparatus to represent a sufficiently wide variety of alternative manifolds. This is a necessary step in being able to represent

[16] This argument was suggested to me by Lewis's explicit statement to the contrary: "[The premise that] a world might possibly consist of two or more completely disconnected spacetimes ... seems to me [to be] ... not a consequence of any interesting general principle about what is possible" (p. 71). Well, it seems to *me* that it is a consequence of a properly understood principle of recombination.

non-physical minds. A first stab at representing non-physical minds might be suggested by a section of the passage quoted earlier, in which Lewis is suggesting some possible extensions of the original Quinean proposal: "and perhaps even ... occupancy of times by non-spatial things" (p. 90). It would be easy enough to represent that possibility within my new framework. Using the superscripts to distinguish dimensions within a single universe, we can represent, for instance, a four-dimensional universe in which some of the properties are located along only the temporal dimension. But this is not enough to represent minds. It is not true to the phenomenological facts about minds. When I put a green book next to a blue book I get a visual experience that contains a green rectangle next to a blue rectangle. One of the rectangles is on the right and one is on the left. One is bigger than the other. These are spatial comparisons alike in kind to comparisons between the books themselves. But if the two colored rectangles are non-physical, then, unlike the books, they do not exist in space. If they are non-physical, then neither of the colored rectangles is to the left of or bigger than either of the physical books. The task is to represent the possibility of objects that do not exist in space—do not bear any spatial relations to the physical books—but still bear spatial relations (or analogues of them) to each other.

My goal here is to allow for the possibility of dualism; I do not intend to argue for or endorse the actuality of dualism. That is a question upon which I shall remain neutral. The one concrete world is very complex, and there are many details that I do not know. For all I know, the one concrete world is dualistic; for all I know, the one concrete world is not dualistic but wholly material.

Let us suppose, for the sake of exploration, that dualism is true. There are, we are supposing, fundamental mental properties that are not physical. What I mean to be supposing is not some weird world with weird mental properties, but a world that is the way our world is if dualism is true here. How do we represent such a world? We can begin by representing all of the physical properties distributed across the manifold in just the way that we are now used to. Of course, we do not really know how to represent this distribution of physical properties, because we don't know enough about what the actual world's distribution is. Let us take care to cling to the dualistic assumption in trying to imagine what this world is like.

Once we have represented the distribution of physical properties, we then have to go on to represent the contents of each person's mind. Let

us consider Jo Schmo. Part of the contents of her mind is already represented, because even if dualism is true, not every mental property is fundamental. Even if some or all phenomenal properties are fundamental, at least some mental properties are fully represented by a functional story about the relations between inputs and outputs and the brain processes that connect them. Let us set those aside. Let us focus on fundamental phenomenal properties.

Begin with the visual phenomena. Here we can think of the mind as a big movie screen, with qualia laid out across the screen. Take the colors on the screen as fundamental properties. Thinking of the screen as a manifold, we can describe the visual content of Jo's mind as fundamental properties distributed across a manifold. The manifold in question has at least two spatial dimensions (I leave it to others to determine whether one's visual phenomenal field has three spatial dimensions to it or only two) plus one temporal dimension. What makes it a temporal dimension, we may allow, is how the properties are distributed across it; I need not suppose here that there is anything intrinsically temporal about a dimension. The properties that are distributed across the temporal dimension of Jo's mind include properties that are not within her mind, since the temporal dimension in question is the same temporal dimension that contains Jo's body and her physical surroundings.

We can represent Jo's "movie screen" with sets assigning properties to locations. When we combine those sets with the ones we used to represent the distribution of physical properties, we have to superscript the points along each dimension so that we can separate the dimensions that are not shared by Jo's internal visual field and the external world. Jo's internal space is not the real external space, even though Jo's internal time is the same time as external time. (Don't confuse her internal time—the time at which certain visual images are appearing to her—with her interpretation of that time. For instance, she may misjudge how much time is passing.)

The set representing the combination of the distribution of physical properties with the distribution of Jo's visual qualia is the very same sort of set that I've used to represent partially connected universes within a single pluriverse. The physical properties distributed across spacetime make up one of the universes, and the phenomenal properties distributed across Jo's three- or four-dimensional movie screen is another universe. The two universes are causally connected along the temporal dimension, since the events that occur in the one are prior to, simultaneous with, or

after events that occur in the other. The two universes are causally connected, since what happens in one influences what happens in the other. Jo often gets a red experience because she looks at something red, and sometimes a red experience in the right relation to other experiences causes Jo to step on the brake pedal.

Proposal: dualistic worlds can be incorporated into my account of the nature of worlds by identifying them with representations of partially connected pluriverses.[17] Let's fill in a few more details.

Jo's auditory experiences form another universe within the pluriverse. This universe is connected temporally and causally to both the physical universe and to Jo's visual phenomenal universe. And likewise for those experiences that we associate with Jo's other senses. Alternatively, perhaps the different kinds of experiences do not comprise separate (though connected) universes, but rather one large phenomenal universe. Are my auditory experiences of the clicks of computer keys in the same space as my visual experiences, or in different spaces? I feel that they are different spaces with important similarities in the patterns of the properties distributed in those spaces, similarities that allow us to correlate the spaces, and that is the assumption I make for the remainder of this chapter, but in reality I leave the question open subject to the reactions of those with more highly developed introspective capabilities than my own. If the auditory and visual universes are separate, then so are the universes that comprise Jo's experiences from the other senses. And likewise, there are separate universes comprised of the experiences of the many other conscious beings that exist in the dualistic world we are considering. And so it turns out that the dualistic world is a pluriverse made up of many many partially connected universes. So, if the actual world is dualistic, we live in a pluriverse.

What if the actual world were physicalistic? The phenomenal properties would not be fundamental mental properties. They would supervene on physical properties. Fully represent the physical world, and we would have represented the phenomenal properties as well. Let us call

[17] If it turns out that causation requires spatial contiguity of cause and effect (as suggested by the worries of n. 14), and it turns out that a dualistic world must represent the concrete world as a pluriverse of the sort I've described, then it will be impossible to have causal relations between non-physical mental states and the physical world. I find this to be a positive aspect to my depiction of a dualistic world, not a challenge to it. My account provides the right relationship between dualism and the continuity requirement for causation.

the physicalistic world "Fizz" and the dualistic world "Doowah". One of these worlds is, by assumption, actual, but we do not know which one. Notice that the distribution of physical properties in Fizz is different from the distribution of physical properties in the dualistic world described above. We do not get Fizz simply by lopping off all of the non-physical universes from Doowah. Jo Schmo cannot tell which of the two worlds she is living in, but that is because of her vast ignorance of the actual world, not because of any great similarity between the physical structures of the two worlds.

If Doowah is actual, then zombie worlds are possible, where a zombie world is just like the actual world physically but lacking in any phenomenal properties. Let us call the world that results from removing the phenomenal properties from Doowah "Zip." In contrast, if Fizz is actual, then zombie worlds are impossible. This is not to deny the possibility of Zip, but only to deny that Zip would be a zombie world—if Fizz is the actual world, then a zombie world would have to be just like Fizz physically, but any world physically like Fizz is also mentally just like Fizz. If one allows for impossible worlds, then there is one that results from removing Fizz's phenomenal properties, but none of Fizz's physical properties. This world is impossible because it will have to represent the same properties as being both instantiated and uninstantiated in the same locations.[18] Let us call the impossible world that is Fizz's zombie world "Imp". It is easy enough to imagine Zip, but our ability to imagine that world does not in any way support the possibility of Imp. Thus, it is not easy to tell whether a zombie world is possible. We must first know which world is actual, Doowah or Fizz, before we can know which world, Zip or Imp, would be a zombie world.

Just as Doowah's physical universe could exist without any of its non-physical universes, so could the non-physical ones exist without the physical one. A world representing a pluriverse containing just those non-physical universes would be a Berkeleyan idealistic world. What about a world like Doowah except that Doowah's physical universe is replaced by Fizz? That world would be a combination of the Berkeleyan world with Fizz. Let us call this combination world Berkzz. That Berkzz is a possible world follows from the principle of recombination. But

[18] It is easy enough to incorporate an account of impossible worlds into my Representationalism. See my "Interpreting Worlds: Identity and Impossibility".

Berkzz is problematic. It seems to contain both Doowah's phenomenal properties and Fizz's phenomenal properties. But if both sets of phenomenal properties can exist as distinct within a single world, then it seems wrong to identify them across worlds—it seems to be a violation of the transitivity of identity. The phenomenal properties in Berkzz's physical universe are identical to the phenomenal properties in Fizz, which are identical to the phenomenal properties in Doowah, which are identical to the phenomenal properties in the non-physical universes in Berkzz, which are not identical to the phenomenal properties in Berkzz's physical universe.

The answer to this worry is that none of these identities hold except relative to high-level interpretations of the properties in the various worlds, and there is no single interpretation that will yield all of these identities. The properties in Fizz are deemed identical to the properties in Doowah in virtue of having the same functional roles. But the other identities are deemed to hold in virtue of sameness of distribution pattern. It is only because Doowah's phenomenal properties are distributed in the same pattern as those in the non-physical parts of Berkzz that we identify them. But the properties in the two worlds do not have the same functional roles. Doowah's phenomenal properties are causally connected to Doowah's physical properties in ways that Berkzz's phenomenal properties are not connected to Berkzz's physical properties. Put in the language of property counterpart theory, the point is that transitivity of counterparthood fails in this case because different respects of similarity are being used in the different counterpart relations that are being appealed to.

A similar challenge from the transitivity of identity arises if we consider a single property F at Fizz that is both physical and mental. It is true to say of F that it could have existed as a physical property that was not mental (for instance, at Doowah), and it is also true to say of that same property F that it could have existed as a mental property that was not physical (again, at Doowah). But this would seem to allow us to say of this one property that it could have existed as two distinct properties. That in itself violates the one–oneness of identity, and, by virtue of that, it will also violate the transitivity of identity. Where "P" names the physical property in Doowah, and "M" names the mental property in Doowah, we seem to have P in Doowah identical to F in Fizz which is itself identical to M in Doowah, despite the fact that P is not identical to M in Doowah.

An answer in terms of property counterpart theory is readily available. All the worry comes to is that F has as counterparts in Doowah two distinct properties—namely M and P. But there is nothing to worry about here, because the counterpart relation is neither one–one nor transitive. This is the sort of situation for which counterpart theory was built.[19] However, I have permitted a higher-level interpretation according to which we can speak of the transworld identity of properties, not merely the counterpart relation. So it seems that I have a worry that other counterpart theorists do not.

But, say I, because the transworld identity permitted by my theory is *just* a higher-level description of the same relationships described by counterpart theory, there should be a straightforward way to translate the counterpart answer to the worry into an answer in terms of identity. And there is. Transworld identity, on my picture, is always relative to a way of conceptualizing the object or property in question. F is identical to M when we are identifying F by its mental role, but not when we are identifying F by its physical role. There is no single way of identifying F that will yield both F = M and F = P.[20]

IV. INVERSION

I should note one oddity about my proposal to identify a mental life with a universe within a pluriverse. I am attempting to accommodate the possibility of irreducible qualia in what is basically a functionalist-style account of possible worlds. My worlds are basically Ramsey sentences for properties in that they are large existential statements that specify roles for every property within the world as a function of the property's distribution among the other properties. As a concrete example of how this oddity plays out, consider the inverted spectrum.

[19] See, for instance, David Lewis, "Counterparts of Persons and Their Bodies", *Journal of Philosophy* (1971), 203–11.

[20] What is at the heart of all of these identity worries is essentialism. In each case my answer to the worry is to remind the reader that the position expounded here is anti-essentialist. If you are an essentialist, then you will think that there is a unique set of properties that is sufficient for transworld identity for any given individual or property. If you are an anti-essentialist you will hold that there is no such unique set; which properties are necessary and sufficient for transworld identity—necessary and sufficient for making it that very same object or property that exists in another world—varies with the conceptualization under which the object or property is being identified.

For the purposes of this discussion I will assume that the actual world is Doowah rather than Fizz. It is plausible that though Jo does not suffer from an inverted spectrum in the actual world, she could have. That would seem to commit us to another world in which all of her color sensations are reversed. But the set that describes her internal movie screen in that other world is the same set that describes her internal movie screen in the (supposed) actual world. The pattern of property distribution is the same in the "two" worlds, and that pattern is all that the set represents (at least at the lower levels of interpretation). So, it turns out that there is really only one world in this story. Which seems to have the consequence that Jo couldn't have had an inverted spectrum after all. What started as an oddity now looks like an objection.

The objection needs to be modified in one respect. If what we want is a world in which only Jo has an inverted spectrum, it is easy to distinguish that one from the actual world. In the actual world, Jo's phenomenal properties are the same as those of the rest of the populace. In a world in which her spectrum is inverted, there will be a difference between her phenomenal properties and everyone else's. Thus, the overall distribution of properties will be different from the distribution in the actual world. To state the objection, then, we want to consider the possibility of everyone having an inverted spectrum, inverted not with respect to their worldmates' spectrums but with respect to the actual world's spectrum. The objection is that the pattern of distribution at the inversion world is the same as that at the actual world. So, they end up being the same world. What looked like two possibilities is only one.

The failure of this objection is best described at the level of property counterpart theory, though we could do it in terms of interpretations of identity instead. Sticking to the lower level of interpretation, saying that Jo and everyone else could have had an inverted spectrum does not commit us to a world in which green experiences are substituted systematically for their actual red experiences (and so on for the other colors). What it commits us to is a world in which a counterpart of their green experience is substituted for their actual red experiences (and so on). (I will ignore talk of people counterparts. We'll have enough to do just getting straight about the property counterparts.) What we need is a world in which counterparts of our green experiences are distributed across our internal movie screens in just the way that our red experiences are in fact distributed across those screens. There is such a world. It is the actual world itself.

A parallel example for objects rather than properties would be; "you and I could have had each other's careers." You could have been a physical duplicate of me, been born and raised in the slums of New York, lost in the fog on Mt. Washington as a teenager, married in Walla Walla, done all the things I did, had all the same beliefs and attitudes I had, and so on, and so on. You could have had exactly my place in the world. And I could have had yours. What makes these modal claims true, according to counterpart theory, is not some world in which someone with my career has your haecceites and someone with your career has mine. Rather it is a world exactly like this one in which the person who has the career that is like my actual career is a counterpart of you and the person who has the career that is like your actual career is a counterpart of me. That is just a world identical to this one. The truth of these modal claims is not grounded on the existence of any other world, but just on the existence of the actual world under an atypical counterpart relation.

Typically my counterpart in the actual world has to be me. I am the object most similar to me. Though there are other worlds in which an object with your career is the one most similar to the actual me (for instance, a world in which that is the only human that exists), in the actual world there is a much better candidate—the object with my actual career. But the "could have switched careers" claim creates a special context. Here we are considering not what I would have been like if certain counterfactual conditions had obtained, but rather what I might have been like—what possibilities are open to me. We attempt to adopt a counterpart relation that is permissive enough to allow the modal assertion to be true. We no longer look for the most similar object to me in the world in question, but just for one that is similar enough. And we try to loosen our enoughness standards to the point where someone with your career can count as similar enough to the actual me, despite the fact that in the world in question (the actual world) there is someone else who is much more similar to the actual me. The effect of loosening our standards is to eliminate all (or almost all)[21] necessary conditions for being me—it is *as if* there were haecceites with no necessary conditions for their instantiation. Depending on other contextual

[21] The assertion "that rock and I could have switched careers" requires even laxer standards.

factors, we often manage to pull it off. In those contexts, we accept the assertion that you and I could have had each other's careers.[22]

Likewise, when the inverted spectrum scenario is proposed we attempt to accommodate the claim of possibility by setting aside all of the causal work that the property of green does for us. We set aside the question of which property is most like the actual world's green in the world in question (the world in question being the actual world itself), instead asking whether any property with the career that red has in the actual world is similar enough to the actual world's green. And we try to adopt enoughness standards that are lax enough to make the modal assertion true. Depending on other contextual factors, we often manage to pull it off. In those contexts, we accept the assertion that green and red could have had each other's careers; we could have had inverted spectrums. It is *as if* green had a haecceites, though it doesn't.

CONCLUSION

The task was to extend the Representationalist account of worlds in order to accommodate possibilities beyond the scope of the original account. One might avoid this task by denying that there are any such possibilities. But here I think the burden is on the person who denies the possibility. The most fundamental modal principle, other than the principle of non-contradiction, is the principle that anything is possible, whatever that means. I say that it means at least this much: if something can be represented consistently in a sufficiently powerful language, then that something is possible. Thus, if we begin with the assumption that pluriverses and minds are possible, the onus is on me to represent them, but once I have managed to represent them, that in itself is evidence of their possibility. It turns out to be easy enough to extend the Representationalist account to accommodate pluriverses, and with pluriverses in hand, we can represent minds as universes within a partially connected pluriverse.[23]

[22] Allowing for these haecceites-like counterpart relations requires a modification of standard counterpart theory. See my "Anti-essentialism and Counterpart Theory", *The Monist* (October 2005).

[23] I thank Juan Comesana, Ben Jarvis, Alyssa Ney, Drew Schroeder, Jim Pryor, Ralph Wedgewood, Alex Byrne, Dean Zimmerman, the audience at Brown University, especially Robert Howell and Jaegwon Kim, and the audience at the Univeristy of Texas.

4. Analyzing Modality

Michael Jubien

1. World Weary

The best known (proposed) *analysis* of the notions of metaphysical possibility and necessity is the surprising and elegant one we owe to the late David Lewis. By the time he put it forward, we were already very much softened up for the general notion that for something to be *possibly true* was for it to be *true* in some *possible world*. So, at the time, it appeared that the analytic project amounted simply to saying what possible worlds are (and tidying up some details). Then we'd have a ''reductive'' analysis of the modal concept, where the reducing primitives were the notion of *truth* and the newly (and finally) elucidated notion of *possible world*. Lewis's elucidation was that possible worlds are infinitely many detached concrete realms, more or less like and including the one we inhabit, but each with its own private spacetime, and each physically inaccessible from each other. (Then he turned his ingenuity to such important details as modality *de re*.)

I believe that any initial plausibility this view might have rests on what may be called a ''fallacy of persuasive terminology'': the decision to call the postulated realms possible worlds, thereby encouraging us to regard them as relevant to nontrivial possibilities—especially possibilities for things that actually exist. I also think any serious effort to legitimize this choice of terminology would undermine any substantial claim of ''reductiveness'' for the proposed analysis. So let's consider these two complaints.

For all we know, there really do exist what we may *neutrally* call detached concrete *realms*. But as soon as we start calling them *possible worlds*, we beg the question of their relevance to our prior notion of possibility. For all we know, there are just two such realms, or twenty-seven, or uncountably many, or even set-many. Suppose there are just a

few, but that all of them happen to include stars. How plausible is it to think that if this is how things really are, then we've just been *wrong* to regard the existence of stars as contingent? Or suppose there happen to be no other concrete realms. Would we happily accept the consequence that we exist necessarily, that we've been overly humble to think we're mere contingent beings? Or imagine, if you will, that there's a distant but very similar planet in our own universe where someone very much like yourself is a playwright and not a philosopher. How plausible would it be to pin the possibility of your having been a playwright on this far-off circumstance? If you find it as implausible as I do, then it should seem all the more implausible if such an individual inhabits an inaccessible physical region instead of a merely distant one. As we'll soon see, I believe the possibility of your having been a playwright has nothing to do with how people are on other planets, whether in our own or in some other realm. It has only to do with you and the relevant property.

But if we've already fallen for the idea that the other realms really are possible worlds, that is, if we're under the spell of the persuasive terminology, then *of course* for something to be possible is for it to be true in (or, more liberally, ''represented as true in'') at least one of them. We were already primed for this by the logical and philosophical efforts of Ruth Barcan Marcus, Saul Kripke, and a few others. But when we just think of the pure, *untitled* ontological picture of detached concrete *realms*, such realms seem to have nothing at all to do with our at-home notion of possibility, even when we imagine there are enough of them to generate intuitively correct truth values if invoked in the spirit of formal modal semantics. Thus it seems to me that the posited concrete realms have a severe and irremediable problem of *relevance*: they cannot reasonably be thought to be relevant to modality as we typically take it, and their irrelevance was merely veiled by the decision to call them possible worlds.

Indeed, when we reflect in this terminologically neutral way, it seems obvious that any other realms that happened to exist would just be scattered parts of the *actual world*, not entire *worlds* at all. They'd be *actually* existing entities. It would just happen that physical reality was fragmented in this remarkable but modally inconsequential way. There would be no call for restricting our notion of actuality to the connected realm we happen to inhabit, nor for viewing the other realms each as ''actual'' with respect to itself but to the exclusion of the others,

nor for viewing individual entities in other realms as modally relevant ''counterparts'' of entities in our realm, nor any of the like.

In ordinary talk, 'the world' sometimes just means the earth. This is a convenient parochial usage. We philosophers, in our grandeur, generally use the term in a much more cosmopolitan way, so as to include everything that exists. Of course it would be just as parochial, only on a far grander scale, to assume that all of physical reality is spatiotemporally connected to us. In fact we simply don't know the extent of the world. Of course, if it happens to be a single, continuous, unfragmented manifold, it doesn't follow that everything true is necessarily true. All the usual possibilities remain, awaiting roundup by a correct analysis.

Thus it seems to me that Lewis's analysis, when stripped of the persuasive terminology, is completely implausible. Suppose this is accepted. We might then wonder whether some further (and plausible) metaphysical assumption might be invoked to render the otherwise irrelevant realms relevant. Thus suppose we go ahead and postulate infinitely many detached concrete realms and suppose we decide to call them possible worlds, *but also* suppose we seek a way to ''oomph them up'' so as to render them relevant to actual possibilities. What would such an oomphing strategy involve?

To enlist a familiar example, right here in what we would then be calling ''the actual world'', we all agree that Humphrey might have won. And according to the strategy, this seemingly local possibility is to be secured by goings on in some other possible world(s) not including Humphrey himself. *Intuitively, this requires the relevant Humphrey-free worlds to have the property of making it possible that Humphrey won*. If we can postulate that other realms exist and have regular properties like being concrete, including people, having presidential elections, and so on, then perhaps we can also find a postulate that would secure that other realms deserve to be viewed *intuitively* as if they had the *modal* property of making it possible that Humphrey won.

One way to achieve this would be to postulate directly that *genuinely* modal properties such as *making it possible that Humphrey won* exist and are instantiated by other concrete realms. Of course this would be a very suspect way of making the other realms relevant, for we could simply have postulated at the outset that Humphrey has such genuinely modal properties as *possibly winning*, thereby saving the elaborate detour through the other concrete realms. Worse, this direct approach would obviously destroy any hope of reductiveness for the analysis and

would create a strong suspicion of circularity as well. Might there be a more promising way to secure relevance?

As Lewis conceived it, the other worlds would have the intuitive modal property because they contained victorious ''counterparts'' of Humphrey. But it's crucial to see that this can't really do the trick unless the counterparts themselves are *already* plausibly seen as having the intuitive property of making it possible that Humphrey won. So, containing a victorious counterpart of Humphrey, as a property of worlds, could do the needed modal work only if it were plausible to view such a state of affairs, once again intuitively, as making it possible that Humphrey won. The most straightforward way of doing this would be to grant *genuine* modal properties, like *making it possible that Humphrey won*, to Humphrey's counterparts, then letting the worlds of the counterparts inherit these properties. But this strategy merely adds an extra level of complexity to the direct approach just dismissed, once again threatening circularity and destroying reductiveness.

There may, of course, be other strategies for providing the right kind of modal oomph for the worlds but without such dire consequences. But here it's important to emphasize that any successful oomphing strategy, with or without ''counterparts'', really must—somehow or other—confer just such intuitive modal properties on the various possible worlds. We must see it as appropriate, given such a strategy, to view other worlds as *making it possible that Humphrey won*. We've already seen that *calling* the other realms possible worlds doesn't achieve this. And we also know that the relations that Lewis officially enlists as ''counterpart'' relations are, *in themselves*, just more or less stringent similarity relations, describable in strictly nonmodal terms. It simply doesn't follow from A's nonmodal similarity to B (no matter how impressive it might be) that A makes something possible *for B*.[1] If someone similar to Humphrey won, that nicely establishes the possibility of someone's winning who is *similar* to Humphrey. But we mustn't confuse this possibility with the intuitively different possibility of *Humphrey's* winning. For the former possibility to establish the latter requires a *further* hypothesis (or presupposition). So when Lewis calls A a *counterpart* of B, we would need to oomph this up so as to entail *both* that A is similar to B (in whatever interesting respects)

[1] In *Naming and Necessity* (Cambridge, MA: Harvard University Press, 1972, 1980), Saul Kripke is apparently thinking along similar lines. See footnote 13, p. 45.

and that some of A's properties establish possibilities *for B*. An apt choice of a similarity relation, in a given context, might indeed make it seem *natural* for us to view A as establishing possibilities for B, but it cannot in itself make it the case that A really achieves this. Any further hypothesis would, in effect, have to confer upon A an intuitive modal property involving B, and hence would have to be modal in its own right. It is for this simple reason that I hold that Lewis's analysis, when bolstered so as to establish the other realms as fully fledged, modally relevant *possible worlds*, cannot be ''reductive'' in the sense of reducing modality to a complex interplay of *nonmodal* concepts. (And for the very same reason, we may now wonder whether we can really regard this strategy as providing even a *nonreductive* analysis of these modal notions, for it appears to rest on concepts of the very sort it seeks to analyze. But I won't press this point.)

I've chosen to begin by discussing Lewis's proposed analysis only because it is so well known. But in fact I think that *any* possible-worlds analysis would be doomed, regardless of the intrinsic natures of the postulated worlds, and for reasons that were tacitly in play in the above comments. Possible worlds analyses are, unfortunately but understandably, encouraged by the formal adequacy of what are sometimes called possible worlds in the semantics of modal logic. But in typical versions,[2] the so-called possible worlds are not assumed to have any specific ontological character. In effect, they're mere *points* or *indices*—they might as well be von Neumann ordinals—and contemporary treatments sometimes do use terms like 'indices' in preference to 'possible worlds'. Any specific modal semantics is of course designed to produce a truth value distribution that reflects some chosen intuitive understanding of the modal operators (among many alternatives). But in no such system does anything at all depend on calling the relevant entities possible worlds, nor on their having any particular internal structure. In modal semantics, sentences are ultimately interpreted with respect to ''worlds'' by means of functions whose arguments are ordered pairs of worlds and nonlogical symbols, and whose values are sets. Proper parts of worlds play no role.

This kind of semantics is of course entirely suitable for its formal role. But its success in this formal arena has no automatic bearing on how

[2] For example, the one given by Kripke in ''Semantical Considerations on Modal Logic'', *Acta Philosophica Fennica* 16 (1963), 83–94.

intuitive modal notions such as that of metaphysical possibility should actually be analyzed. Despite this, many philosophers seem to think that developments in pure logic have provided the foundation for the proper analysis of our modal notions—a foundation that includes ''possible worlds'' as an indispensable part.

There is an important irony here. For when *philosophers* appeal to possible worlds in discussing modality, *their* worlds are anything but mere indices. Instead they're stunningly complex, ''maximal'' entities of one ontological variety or another. Then these ''worlds'', collectively *and* as a result of their internal complexity, are exploited in trying to analyze (or at least ''capture'') possibility and necessity.

But really, if something is possible, for example your having been a playwright instead of a philosopher, then of course that is a truth right here in the actual world. It's a fact about how this very world might have been, not a fact about how some *other* world is. It's very odd to think that any truth that doesn't obviously concern ''other worlds'' should either *be* a fact about other worlds or else *depend on* facts about other worlds. The other worlds, after all, are *not* the actual world.

I think typical modal truths are just facts about our world, and generally facts about very small parts of it, not facts about some infinitude of complex, maximal entities. At the very least, if there's a plausible way of understanding modal facts without going beyond what is actual, then it ought to be preferred to any account that makes them depend on (or be identical with) other-worldly matters. Thus I will later strive to analyze these modal concepts in a way that appeals only to what actually exists.

These recent considerations reflect a deep and fundamental weirdness that I believe must confound *any* proposed possible-worlds analysis of modality. I think this weirdness has generally managed to escape our notice simply because we've somehow fallen deeply and uncritically under the catechistic spell of possible-worlds talk. But when the weirdness is recognized, one can't avoid the queasy feeling that something very basic has gone awry.

In a nutshell, the weirdness is this. Suppose it's necessary that all A's are B's. This is supposed to mean that in every possible world, all A's are B's. So the necessity arises from what goes on in all the worlds taken together. There's nothing about any world individually, even in all of its maximal glory, that forces all of *its* A's to be B's. It's as if it just *happens* in each world that all of its A's are B's, that from the

strictly internal point of view of any world, it's *contingent*, a mere coincidence. But then shouldn't we expect that this internal contingency will not be repeated in every world, that there will be worlds where some A's ''happen'' not to be B's? After all, nothing within any given world prevents it, and these are supposed to be *all* the possible worlds.

The fundamental problem is that in world theory, what passes for ''necessity'' is in effect just a bunch of parallel ''contingencies''. The theory provides no basis for understanding why these contingencies repeat unremittingly across the board (while others do not). As a result, it provides no genuine analysis of necessity. What looks superficially like an analysis is really just a trivial consequence of the intuitive notion: *of course* if something is necessary, *and* there are all these ''possible worlds'', then the something that is necessary will be true in each of them. But that doesn't tell us what forces it to be true in each of them, in other words, what its necessity consists in.

If you ask a possible-worlds theorist whether there are any worlds in which horses aren't animals, you will be told that there are not. And if you ask how we know this, the answer will be that the necessity of horses' being animals is intuitively clear. I think this response should immediately raise the suspicion *that the necessity doesn't arise from how the worlds are, but rather that the worlds are taken to be as they are in order to capture the intuitive necessity*. Yes, it does seem intuitively clear that horses are animals of necessity, but now exactly what is the source of this clear intuition? Is it reflection about an infinitude of complex, maximal entities, or is it reflection about the properties of *being a horse* and *being an animal*? A possible-worlds analysis seems to require intuitive insight into the former since the necessity is supposed to consist in—to emphasize, just *be*—this fact about all possible worlds. We would somehow have to intuit the pattern of locally accidental instantiation of those properties throughout all possible worlds. Just thinking about the properties themselves could not do the trick.

But in fact it's *precisely* what does the trick. We try to imagine a horse that isn't an animal and we fail. In other words, we try to imagine something instantiating *being a horse* but not instantiating *being an animal*, and we fail. Surely this isn't about infinitely many complex, maximal entities, but rather just about the two properties. *I find it a stunning irony that it's on the basis of this very spare thought experiment, not involving ''possible worlds'' at all, that people so readily declare that there's no possible world in which a horse is not an animal.*

The worlds, to the extent that speaking of worlds might be reasonable, merely reflect our thoughts about the properties. I claim that these thoughts are thoughts about the real source of possibility and necessity. The possible worlds of logicians and philosophers are, respectively and at best, convenient mathematical and metaphorical devices. The former device distributes truth values to modal sentences the way the logician wants them distributed. The latter, however elaborated, captures the philosopher's intuitions about relations among various properties. But if we really want to *analyze* possibility and necessity, we will have to look past the veil of possible worlds to the real source of these modal notions.

Thus I ask you, for the occasion at hand, to set aside any preconceptions you may have about how modality should be analyzed, including especially the powerful prejudice that any analysis must somehow involve special entities called possible worlds, whether of Lewis's variety or any other. For I'm going to point toward a very different analysis, one that I find immensely more plausible metaphysically. In a certain very general way it's similar to Lewis's analysis, for each has its own special ontology, and the items of each ontology have a built-in ''modal character''. Both analyses also rely on further properties or relations.

Lewis's ontology was the infinitude of detached concrete realms, but it seemed to require the added assumption of extrinsic modal oomph, and the further properties and relations were the property of *truth* and an uncertain number of *counterpart* relations, of course also endowed with that special modal oomph.

In contrast, I believe there's just *one* world, which I don't mind calling the actual world. But I don't believe it's entirely concrete, and I have no opinion as to whether the concrete part is spread out in a continuous single spacetime, or fragmented into who knows how many detached and mutually inaccessible parts. But, as you already know, I don't think these unknown circumstances have any bearing on the analysis of modality, so I think our ignorance here doesn't matter.

I think modality has to do with relations involving the *abstract* part of the world, specifically with relations among (Platonic) properties.[3] Although I don't think it's ultimately essential, this view of modality is

[3] Ted Sider has suggested that this might be called a ''governance'' view, and in a certain way this is entirely apt. It is a view according to which modal facts in the concrete part of the world are ''governed'' by relations that hold in the abstract part.

most easily described if we posit an *abundant* realm of properties. Many philosophers, going back at least to Plato, have exploited properties for a wide variety of important philosophical objectives quite apart from coping with modality. So I think it's a dramatic understatement to say that Platonic properties have broader and deeper antecedent philosophical credibility than do modally enhanced detached concrete realms. As putative entities whose existence we can't verify in any remotely direct way, properties are overwhelmingly more plausible.

2. UNTIL THE REAL THING COMES ALONG

The work of such philosophers as Quine, Marcus, and Kripke amply illustrates that the topic of necessity and possibility is inevitably bound up with the general topic of reference. We cannot expect to arrive at a plausible understanding of modality without a simultaneous, if not prior, understanding of how our singular terms and quantifier expressions are connected to individual entities in the world. Nowadays it is often claimed, for example, that ordinary proper names ''refer directly'' to individual things and that (typical) quantifier expressions ''range over'' individual things. But it's a surprisingly delicate matter to come to a clear understanding of the word 'things' (or 'objects' or entities', etc.) as it occurs in statements like these. In fact the matter is so delicate that it's a little difficult to discuss it clearly (as we may soon see!). This delicacy may partly explain why the topic has received virtually no attention: it's so delicate that it just goes unnoticed. Typically, it is *taken for granted* that we understand these terms, even, for example, in debates about whether proper names ''refer directly to *things*''. But I think it's a mistake to take it for granted, and that we have to resolve the matter before we can arrive at a plausible analysis of the modal notions (or a plausible theory of reference, for that matter).

In an effort to illustrate, let's briefly consider the intuitive proposition that all horses are animals. The sentence 'All horses are animals' evidently expresses this proposition, but there are two different ways to think about how it achieves this. One is to take the phrase 'all horses' as somehow picking out (or ranging over) the world's *horses,* in one fell swoop. Then the predicate asserts of these horses that they are animals. Seen this way, the phrase 'all horses' is a ''horse-quantifier'' and the sentence is only ''about'' horses. There may be various ways of

developing this approach in detail. It is probably the one that best reflects our intuitive understanding of the English, but unfortunately it seems to require a plenitude of quantifiers, matching the plenitude of categories into which things may fall.

The other way, which is reflected in first-order logic, is to take the phrase apart, roughly seeing 'all' as (in general) ranging over lots of things in addition to horses, but then letting 'horses' restrict the application of the subsequent predicate to the things in the wider range that satisfy the condition of *being a horse*. The effect of this approach is to make the sentence assert that each and every *thing* (under consideration) has the (''conditional'') property of being such that *if it's a horse, then it's an animal*. On this ''unitary'' view of quantification, the sentence is just as much ''about'' (say) *statues* as it is about *horses*. It's about each and every *thing* (in the relevant universe of discourse), irrespective of any familiar and narrower category into which it might fall. Thus we get by with a single, category-neutral quantifier.

These two treatments pair naturally (though maybe not inevitably) with contrasting ways of thinking about (and describing) our physical surroundings, and although it's a little tricky to convey this clearly, I'll try. In parallel with the multiple-quantifier view, we may see physical reality as made up of *horses*, *statues*, *stars*, *puddles*, and so on. Of course each horse (statue, etc.) also has the property of *being a physical object*, but on this way of thinking its status as a horse is somehow ''more fundamental'' (here is where clarity may seem to slip, but bear with me). The picture that pairs with the single-quantifier view reverses this. We see physical reality as made up of what are ''fundamentally'' just (*physical*) *things* or *objects*. Some of these things of course have the property of *being a horse*, others the property of *being a statue*, and so on.

So what? Aren't the two pictures essentially the same? Is there any real difference between viewing physical reality as consisting of horses, statues, and so on, and viewing it as consisting of physical things that *are* horses, statues, and so on? Surely it's easy to see the latter as just a slightly wordier equivalent of the former and I think that, properly understood, it is. But the fact is that we speak and reason differently about horses and statues from the way we do about the very physical objects that in fact are horses and statues. Let me try to clarify this dark-sounding claim.

Imagine that we're thinking of physical reality as composed ''fundamentally'' of horses and statues, and so on (which of course *are* physical

entities). Now consider a statue. On this view, when we refer to or quantify over this entity, we are referring to or quantifying over *a statue*. I think this has a profound effect on which properties we're inclined to regard as *essential* to the entity, the most obvious being its status as a statue. Thus, merely for example, a possible-worlds enthusiast would very naturally be inclined to think that if we consider this entity ''in another world'', we will automatically be considering a *statue* (in fact, *that* statue).

But now shift to the view of physical reality as composed ''fundamentally'' of physical objects. On this view, when we refer to or quantify over this very same entity, we're referring to or quantifying over a *physical object*, one that indeed *happens* to be configured as a statue, but very easily might not have been, might have been a birdbath or a planter, or just a blob of clay of no significant shape or artifactual status. It won't be seen as essential to the mere physical object that it's a statue. And the fan of possible worlds will think that if we consider this entity in other worlds, then yes, it will *not* automatically be a statue. In some worlds it will be a birdbath or a blob.

I believe that this example illustrates a deep-seated (and far-reaching) tendency to think differently about physical entities, depending on whether we think of them as objects of familiar kinds (statues, horses) or think of them merely as physical objects. I have called this tendency in our thinking ''the great divide'', and I believe its recognition is the key to the solution of a variety of puzzles (and also to the resolution of some important issues) in metaphysics and philosophy of language.[4] But for the moment, let's just consider this. If something along these lines is correct, and the specific example is accepted, then it will obviously be very dangerous to move in our modal thinking from claims that are, intuitively, ''about the statue'', to claims that are, again intuitively, just ''about the physical object''.[5] It won't follow from our conviction that the statue is essentially a statue that the physical object is as well. (For it is also our strong conviction that the physical object might not have been a statue.) It will emerge that I believe there's a clear and straightforward way to reconcile such seemingly incompatible

[4] For a detailed discussion, see ''Thinking About Things'', *Philosophical Perspectives* 15 (2001), 1–15.

[5] I call the error of drawing conclusions ''about'' the *physical object* from claims ''about'' the *statue*, ''object fixation''. For a full discussion, see ''Thinking About Things''.

claims, and to do so without resorting to the implausible view that they're about distinct entities. The reconciliation depends on understanding these claims not merely as claims about certain *entities*, but also as claims about certain (different) *properties*. This strategy will play a vital role in the general approach to modality taken here, and will be treated in more detail soon.

I hope we now have a better sense of what it might mean to distinguish between taking our physical surroundings as consisting fundamentally of (e.g.) horses and statues *versus* taking it to consist fundamentally of physical objects (that happen to fall into various familiar categories). In what follows I adopt the latter picture, and with it the single-quantifier approach, thereby siding with first-order logic as opposed to intuitive English. I do this for four main reasons. First, it is fully general. It readily accommodates physical objects that fall into no familiar categories at all (and the alternative may not). Of course it achieves this without losing the horses, which are simply physical objects that have a certain familiar and important property. Secondly, it is more ''naturalistic''—it doesn't promote categories that happen to be important to us (horses, statues) to any special status. It's an approach that could be happily adopted by human and Martian alike. Thirdly, it does not appear to prejudge any interesting questions about how to analyze modality. For instance, it is fully compatible with a possible-worlds approach. Fourthly, and most importantly, there is a very important sense in which *being a physical object* really *is* our most fundamental category for dealing with the external world: it's the only natural property that everything out there has. The other categories are all more restricted and hence appropriately viewed as special cases of this fully general and fundamental category. Thus I am going to aim for an analysis in which our (singular objectual) quantifiers are taken to range over all physical objects, not just over those that happen to fall into familiar, narrower categories, and, what is more important, not under any requirement that there be any unique, nontrivial privileged category (whether familiar or unfamiliar) associated with each such object in that range.

The analysis I have in mind depends heavily on certain special sorts of properties, which I will now begin to describe. First, for any entity of any sort at all, abstract or concrete, I assume there's a property of *being that specific entity*. When it comes to physical entities, I accept Quine's very liberal and naturalistic account: a physical object, or thing, is just

the material content—the physical stuff—of any region of spacetime ''however disconnected and gerrymandered''.[6] This view invites us to take physical stuff as a primitive concept and to conclude that the property of *being a specific physical object* is identical with the property of *being such and such specific physical stuff*, the stuff of which the object is (intuitively) composed. I am going to view physical objects in just this way for the remainder of the discussion, though I'll mention a serious alternative (actually more of a special case) a bit later. Some consequences of adopting the Quinean view will surface soon.

Properties of this general sort—*being such and such specific entity*—are often called *haecceities*, but this term sometimes carries connotations I don't favor—notably non-qualitativeness—so I'll avoid it. Instead I'll just rely on the intuitive force of the notion itself and call these properties *entity-essences*.[7] When the entities in question are physical objects, I'll call them *object-essences*. The key feature of an entity-essence is that it could not be instantiated by any entity other than the one that in fact instantiates it. To this extent, entity-essences have a certain built-in modal character, about which I'll soon say more.

Entity-essences are special cases of a much more general notion. In the domain of the physical, this more general notion encompasses not only object-essences, but also *dog*-essences, *planet*-essences, and so on. For any kind k and specific entity of kind k, that entity instantiates what I call a *k-essence*. To illustrate, suppose you have a dog. Your dog, and of course any dog, is a *specific* dog. So there's a property of *being that specific dog*. This particular dog-essence is in fact instantiated by a specific physical object, which therefore of course also instantiates a certain *object*-essence.

We may now ask whether the dog-essence is identical with the object-essence. The extremely important answer is that it's *not*. They're two different properties, and here's why. As previously remarked, the object-essence could only have been instantiated by the specific *thing*—in other words, under the Quinean conception, the specific *matter* or *stuff*—that does instantiate it. But the dog-essence, which is in effect just our (Fregean) concept of the specific dog, allows that different stuff might have instantiated it. It's uncontroversial that some of the

[6] *Word and Object* (Cambridge, MA: MIT Press, 1960), 171.

[7] Those who don't feel any intuitive force are encouraged to go along anyway. The alternative to be described later may be more appealing.

stuff that is in fact incorporated into the dog's body might not have been; for example, if its nails had just been clipped. In this event, the object instantiating that very dog-essence would have instantiated a *different* object-essence, for it would have been a different physical object.[8]

Dog-essences, and the like, thus also have a built-in modal character, as promised. Our concept ''of a certain dog'' (or planet, table, etc.) tolerates a range of possible constitutional differences, as we just saw, and other differences as well. We ordinarily allow that at least somewhat different stuff might have constituted ''the same dog'', and that ''the same dog'' could have been smarter or better trained or born a little later or earlier, or in a different place, and so on. I think this is just a matter of empirical fact about how we actually think about ''specific dogs'', and so on. Exactly how much constitutional difference we tolerate when it comes to a certain dog (or planet, etc.) is unclear, and I'll make no claim about it. All that matters is the unquestionable fact that we do tolerate some. Any uncertainty about how much just reflects the vagueness of nearly all familiar general terms, and addressing that phenomenon is a different (and of course considerable) philosophical project. For present purposes, it's best if we pretend there's no vagueness. I think that nothing I'm going to say here would be undermined by any specific account of, or resolution of, vagueness.

Now, it may seem that the present notion of object-essence commits us to the oft-maligned doctrine of mereological essentialism. I think it actually doesn't, though I believe we are so committed. The real source is what I take to be the natural reading of the Quinean conception of physical objects. *If* to be a physical object is just to be the material stuff of some region of spacetime, then, on this reading, unless we have *that very stuff*, we don't have *that physical object*. Considered in the abstract, it is very difficult to deny this naturalistic view. It is also very difficult to conjure up a contrary view to put in its place. For a contrary view would be one in which different stuff could constitute the same physical object. But such a view would have to include a *general* account of the conditions under which such modal ''metamorphoses''

[8] This makes the term 'k-essence' rather misleading, for in general a k-essence is not an intuitive ''essence'' *of a single thing*. It's an ''essence'' we would *intuitively* associate with (say) a single *dog*, but now we need to be mindful of the fact that different *things* could have had that same dog-essence. Greg Fitch and Ted Sider have criticized this terminology for this reason, but I haven't been able to think of an alternative that relieves the intuitive discomfort without being overly cumbersome. (One does get used to it.)

are tolerated, and no such general account has ever been articulated. The Quinean account, on the other hand, is clear and simple, and has been well articulated. As long as we may accommodate our considered intuitions under this account, I believe we have no serious alternative but to accept it.

Given this account, we may do a simple ''essentialist''-style thought experiment that makes mereological essentialism seem entirely mandatory. First, imagine some *arbitrary* physical stuff. You're not allowed to know whether it's a dog or a car, or the like, or a widely scattered entity of no familiar kind at all. (After all, the principle under consideration is a fully general, abstract principle about *things,* not a principle about things of familiar kinds per se.) Now try to imagine a situation in which that very stuff exists, but without some of it existing. I believe any sincere and careful effort to achieve this must fail. The best we can do is to imagine *some* of the original stuff existing without the *rest* of it existing. But this is simply not to imagine the original stuff existing.

This reflection has a striking consequence under the present conception. Since no stuff could have existed without all of it existing, no physical object could have existed without all of its parts existing, for each of its parts is just some of its stuff. (Of course, none of this is to say that the stuff—that is, the thing—couldn't have been configured differently.)

I believe that this thought experiment establishes mereological essentialism (for physical objects) as firmly as any thought experiment could establish any nontrivial metaphysical thesis at all (though of course it does so only in company with the present understanding of the Quinean conception of physical things, which is needed to support the transition from the stuff to the thing).

Someone who rejects this thought experiment (while retaining Quine's conception of physical objects) is, I believe, insisting on a fundamentally mysterious conception of the parts of things. On this conception a thing may, under unspecified conditions, (modally) gain or lose parts while retaining its ''identity''. But it's very difficult to entertain such a view coherently without relying covertly on mereological essentialism. For suppose someone claims that an actual thing, X, with actual proper part Y, could have failed to have Y as a part. What could support such a claim? Is it that we can imagine the actual thing X-Y (call it ''Z'') existing, and then somehow think of it as being X? This can't work. For if an arbitrary thing like X is supposed to be capable of having different parts, then its parts, notably Y and Z, must also be

capable of having different parts. Thus how do we know that the *actual* thing Z, in the *imagined* situation, wouldn't have Y as a part? The only obvious and general way to ensure that it wouldn't would be covertly to assume that in the imagined situation both the actual Z and the actual Y would have the parts they *actually* have! In other words, the only clear and completely general way of achieving the desired outcome would be to presuppose mereological essentialism! (Of course for the opponent of mereological essentialism, the fact that Y isn't *actually* a part of Z is officially irrelevant, and calling Z ''X-Y'' in the *imagined* situation would simply beg the question.)

So why have so many people rejected the thesis? The answer is that they're thinking of objects as if they were ''fundamentally'' of certain familiar *kinds*—as discussed above—not just of objects qua physical objects. They're thinking that the fact that (for example) a *car* could have lacked some of its actual parts shows that a *physical object* could have lacked some of its actual parts.

But it doesn't show this at all, and the reasoning illustrates the above-mentioned error of *object fixation* (which, in lighter moments, I also call *property repression*). We fall into this error when we reason from the fact that an object of a certain explicit *kind* has a given feature to the conclusion that a *physical object* has that same feature. Very roughly, the problem is that the original statement is partly about the *kind* and the inferred conclusion is not. So it's object fixation in the sense that we fixate on the relevant *object* and repress the role of the relevant *kind*. Often such inferences are innocent enough, but not always, and not in the cases typically offered as counterexamples to the thesis.

From the present perspective, here is what's going on with the car. If, say, your car's rear bumper had been stolen in the night and secreted away to a distant location, never to be reattached, in the morning a different (and somewhat smaller) physical object would have instantiated the same car-essence that your car actually instantiates. (And if you decided to replace the bumper, then yet a third physical object would later come to instantiate that same car-essence.) *But no physical object would have lost any parts*. The object that had been your car would still exist, but would now be grossly scattered and no longer instantiate any car-essence. And the bumper-deficient object that would then be your car—that would then instantiate *the same* car-essence—never included the bumper as a part in the first place. When the bumper was attached, the part of your car not including it was just a proper part of a car, and so

did not instantiate any car-essence. In short, when an object of a familiar kind, like a car, undergoes an intuitive change of parts, what really happens is that a different physical object comes to instantiate the relevant k-essence.

Clearly this diagnosis is fully general. So mereological essentialism is unaffected by the usual alleged counter-examples, and we don't have to abandon our everyday beliefs about the parts of cars and dogs in order to accept it as a consequence of the present naturalistic conception of physical objects. That's good news because the naive "counter-example case" against this doctrine has always seemed somehow too cheap and easy. Now we have what I hope is a good inkling why: the inference in the counter-example case jumps illegitimately across the great divide in how we think about physical objects.

One might evade this happy conclusion by denying the present account and putting a different one in its place, one that left the parts of objects "modally negotiable" so as to conform to our everyday intuitive views about things like cars and horses, and so on. But, as suggested above, this is a very daunting project. The reason is that any completely general theory must settle which parts of *arbitrary* things are essential to them. But we have very conservative intuitive opinions with things of some *kinds* (e.g., a Stradivarius) and very liberal ones with others (e.g., the 12th green at Augusta National). So no general prescription is going to work across the board. Our theorist would be forced, in effect, to tell us kind by kind. Thus: dogs, planets, cyclotrons, handles of cheese graters, and so forth.

But it's actually worse than this. For often a physical object of one kind could have been configured so as to be of a different kind, and with different intuitive standards of part-essentiality between the two kinds (perhaps, e.g., a Rodin versus a birdbath). The varying vagueness of general terms thus seriously undermines the project of giving an anti-mereological-essentialist account of physical objects, and the prospects for a fully general account are accordingly very dim.

It's a major advantage of the present account that it leaves the vagueness exactly where it belongs—with the general terms—while incurring no obligation to remedy it. And so we have yet another consideration that favors the decision to take our quantifiers as ranging over physical objects independently of (further) category. For then we have a standard of part-essentiality that holds with complete uniformity, and we avoid the problem of stating different standards for different

categories of entities. This approach is of course compatible with widely varying intuitions regarding dogs, rare violins, famous greens, and so forth. The variations will of course affect dog-essences, Stradivarius-essences, and the like, and appropriately so, but without affecting object-essences.

So far we have adopted a specific account of physical objects, we have postulated an ontology of Platonic properties, resolving to analyze modal notions in terms of them, and we've considered a few specific sorts of properties that are going to play a large role in the effort. Further, we've just seen how these properties help reconcile an important metaphysical doctrine with our ordinary thinking about the world. Next we employ these same assumptions and tools in thinking briefly about the vital topic of proper names.

3. IT DON'T MEAN A THING

Saul Kripke says that names are ''rigid designators''.[9] Ruth Marcus speaks of proper names as ''identifying tags'', devoid of meaning.[10] These powerful voices have been heard, inspiring the contemporary ''direct reference'' view according to which proper names refer directly to specific entities and have no descriptive content. Sentences containing names are seen as ''loading'' the entities referred to by the names ''directly'' into the intuitive propositions they express. This view has become one of the most widely accepted doctrines of contemporary philosophy.

But, from the present perspective, it cannot be right. It is an indisputable modal fact that (the planet) Venus could have had some different parts. If the name 'Venus' simply contributed a physical object to claims expressed with the help of that name, it would follow that some physical object could have had different parts. But we have apparently just seen that no physical object could have had different parts. And all we needed in order to see it was a seemingly innocent and naturalistic conception of physical objects, and a seemingly incontrovertible thought experiment about physical stuff.[11] The falsity of the direct

[9] *Naming and Necessity*, 48 *passim*.

[10] In ''Modalities and Intensional Languages'', reprinted in *Modalities* (New York and Oxford: Oxford University Press, 1993), 11.

[11] Indeed, a thought experiment that happens to be very Kripkean in spirit.

reference view is a very important consequence of these seemingly innocent and straightforward metaphysical considerations.[12] (At the very least, a proponent of that view will have to find a way to undermine such considerations before the view can continue to seem plausible.)

So the name 'Venus' does not simply contribute a physical object to the propositions expressed by sentences containing it. I believe there is no serious alternative but to conclude that it contributes a *property*, the property of *being Venus*. Different physical objects could have instantiated this property since it is undeniable that (whatever the semantical details may be) 'Venus could have had different parts' expresses a true proposition. So not only does a name not *simply* contribute a physical object to the proposition, it doesn't contribute one at all. Its function can only be to contribute whatever property the name expresses.

Although there isn't space to defend the view here, I believe that every ordinary proper name expresses a unique k-essence. (In the case of Venus, the relevant kind is *celestial body*.) Typical names express person-essences, dog-essences, ship-essences, and so forth. That ordinary proper names express k-essences may now be added to our growing list of assumptions.[13]

Kripke has his own seemingly incontrovertible argument in favor of rigid designation, so let's review it. He urges, in effect, that when we utter a sentence like ''Humphrey might have won the election'', we're surely speaking about the very same person we speak of when we describe the actual situation by saying ''Humphrey lost''. From this he *infers* that the name 'Humphrey', when it occurs in a modal context, designates *the same thing* it designates when it occurs in a nonmodal context.

There can be no doubt that we're talking about *the same person* in the two sorts of contexts. But—from the current perspective—it doesn't follow that we're talking about *the same thing*. The (very natural) tendency to draw this inference is another example of what I called object fixation above. It's the same mistake we found in naive arguments against mereological essentialism. (A *car* could have had different parts, so a *thing* could have had different parts.) In moving from the

[12] A more thorough discussion may be found in ''Thinking About Things''.

[13] A full defense of this view involves a careful consideration of the different ways proper names enter the language. Orthodox views tend to go astray partly as a result of focusing only on common or idealized cases.

person to the thing, we jump the great divide, suppressing the role of the person-essence, *being Humphrey*, in the original proposition, imagining instead that it merely concerns some specific *thing*. But we have just urged that names do not refer directly to specific things. As a result, again from the present perspective, there is really only one plausible reading of sentences containing names.

'Humphrey lost' means that something instantiates both *being Humphrey* and *losing*. 'Humphrey might have won' means it is possible for something to have instantiated both *being Humphrey* and *winning*.[14] These two propositions *are* ''about the same person'' because to instantiate *being Humphrey* is to instantiate a *person-essence*—the property of being a certain specific person. But the truth of the two propositions clearly does not require that the thing that instantiates *being Humphrey* and *losing* might *itself* have instantiated *being Humphrey* and *winning*.

Thus no proposition ''about Humphrey'', modal or otherwise, (thereby) contains any specific physical thing as an intuitive constituent. The proposition *that Humphrey lost* is just *not* a ''singular proposition''. It's the *general* proposition *that something (or other) instantiates both being Humphrey and losing*. The feeling of singularity arises from the fact that *being Humphrey* is a property of a sort that can have at most one instance—a ''singulary'' property, if you like. Because we normally think of Humphrey as a specific person, object fixation leads us to overlook the possibility that a different thing might have had the property of *being Humphrey*. Then this mistake results in the erroneous assumption that the proposition includes the *actual* instance of the property as a constituent. (Of course the existence of a specific instance is needed for the proposition to be true. But this is just the familiar royal road of quantification and truth.)

It is important to emphasize that in rejecting the direct reference view, rejecting (typical) ''singular propositions'', and labeling certain ordinary inferences as fallacious instances of object fixation, and so on, I am relying heavily on something approximating the present view of physical objects. But I believe *any* plausible view would have the same consequences (and I promised to mention an alternative below). Thus the burden, I claim, is on the shoulders of direct reference theorists:

[14] Of course we will see that this isn't the ultimate analysis. The modal term will have to disappear.

they owe us an account that legitimizes ''object fixation'', overthrows mereological essentialism, and provides a *general* standard of part essentiality in conformity with our intuitions about arbitrary objects of different kinds. This is a very heavy burden given the simplicity and straightforwardness of the present alternative.

We have now seen the key entities in the special ontology of the current view of modality, and we've also seen how these properties help reconcile an important metaphysical doctrine with our ordinary thinking about the world. We may now proceed to the analysis of modality. It depends on these properties, and also on a certain important relation among them: a relation I hazard to call *entailment*.

4. MY TRUE STORY

We all agree that *being square* ''entails'' (or ''involves'') *having linear sides*; that *being yellow* entails *being colored*; that *being a wife* entails *being married*; that *being a horse* entails *being an animal*; and so on.

The idea of entailment between properties is not new. But I believe it's generally misunderstood, and dramatically so, and even by philosophers who are entirely comfortable with Platonic properties. These philosophers typically offer a modal analysis of the notion, specifically: for P to *entail* Q is for the proposition that all P's are Q's to be necessary. Then they generally take *necessity* to be ''truth in all possible worlds'' (or else take it as primitive).

As we shall see, I think this analysis is completely backwards, that the necessity rests on the entailment, not vice versa, and of course that ''possible worlds'' play no role at all in the matter.

At this stage it may be helpful to recall some basic Platonic principles. We are taking properties to be genuinely existing entities, and all entities of course stand in various relations to each other. We are also taking it that for something to be a horse *just is* for it to instantiate the property of *being a horse*, and similarly for being an animal. But this venerable Platonic principle would lack plausibility *unless* properties were taken to differ from each other in their intrinsic natures (which of course are atemporal and immutable on this ancient view). From this perspective there's an utterly natural explanation of our intuitive feeling that any horse *must* be an animal. It's that the two properties' intrinsic natures together guarantee it. We may therefore see this

connection as an ''intrinsic relation''—one that holds between the two properties strictly as a result of their individual intrinsic natures.

It is therefore an *ontological* feature of this Platonic view that it's *automatic* that if and when something instantiates *being a horse*, it also instantiates *being an animal*. This is why the proposition that all horses are animals would be true ''in each possible world''. It wouldn't ''just happen'' in each world that all the horses were animals (the way it would just happen in a given world that all the horses were wild). Something about *being a horse* would make it so, whereas nothing about *being a horse* would make a wild horse wild. I call that something—that intrinsic relation—*entailment*. Thus I hold that the necessity of the proposition consists in the entailment of the one property by the other (and of course that ''possible worlds'' of any variety are superfluous).

I have no opinion about the ultimate nature of property entailment (though in the past I have entertained certain hypotheses). In fact I think the ultimate nature of entailment is unlikely to be discovered by philosophical inquiry. For practical purposes, we may as well think of it as primitive (but with no commitment to simplicity). Although our ignorance here might be disappointing, in no way does it undermine the authority of the conclusion that entailment does indeed hold or fail to hold between pairs of properties as a result of their intrinsic natures. And this is all we need in order to analyze modality. We already have a big chunk of the analysis. For example, the necessity of the proposition that all horses are animals consists in the fact that the property of *being a horse* entails the property of *being an animal*.

Now is a good moment to pause and address a touchy question. Is this analysis ''reductive''? Just what is a reductive analysis in the first place? As I see it, an analysis of a concept tells us what the concept *is* by telling us what its constituents are and how they are combined in the concept. So if the concept under analysis has a certain characteristic feature, say being a *mental* or a *mathematical* concept, then that feature must *also* somehow be present in the analysis, or else the analysis cannot be correct. *De dicto* necessity is generally taken to be a paradigmatic *modal* concept. If it is, and if the correct analysis of the necessity of all A's being B's is—let's just say schematically—that the property of *being (an) A' bears relation R to the property of being (a) B'*, then it seems to me that the relation R *must* be modal—even if we might not previously have thought so—for R would then be the source of this species of necessity. The broader moral is that *any* concept has the

characteristic feature F if and only if a correct analysis of that concept also has F. Of course one could become convinced of the correctness of an analysis and then conclude that the concept did not have the originally supposed feature F after all. Alternatively, one could conclude that what initially seemed a congeries of non-F concepts was really imbued with F after all. We might have it either way. What we cannot have is a correct analysis of an F-concept in entirely non-F terms.

Of course I have taken the schematic analyzing relation R to be entailment in the sense described above: a standing relation that holds immutably and intrinsically between certain pairs of properties. On this view it is the source of necessities like the one just considered and therefore a constituent of any parallel claim of *de dicto* necessity. So, to the extent that such necessities are *modal*, so must be this relation, even if one might initially have thought that standing intrinsic relations between Platonic properties are unlikely to have such a feature. I thus claim no spectacular conceptual reduction in offering the present analysis. I think no such reduction could possibly accompany any correct analysis of any concept.[15]

Now let's recall the earlier discussion of Lewis's proposed analysis, which is often thought to be importantly ''reductive''. I argued that when we think of Lewis's ''worlds'' merely as detached concrete realms, they have no relevance to nontrivial possibility, and the analysis, understood in these terms, would be implausible. We are now in a position to see that it would be implausible *precisely* because it would be ''reductive'' *in the sense of purporting to analyze a modal concept with only nonmodal resources*. On the other hand, if we view the ''worlds'' in a way that supplies the modal oomph required for them to be modally relevant, then the result ceases to be reductive in this sense and, if the oomph is understood in the most natural and direct way, it even threatens circularity. Of course, there might be other ways of understanding it that were noncircular and even ''reductive'' in some more plausible sense.

For example, there is a sense in which Lewis's proposal—modal oomph aside—would be genuinely ''reductive'' simply as a result of its structure, and this may help explain the widespread impression of reductiveness. Lewis often expresses it by saying the modal operators

[15] I thank Greg Fitch and Karen Bennett for insisting that I say more about reduction and, in particular, about the modal status of the concept of entailment.

are quantifiers over worlds. For example he says this in discussing the box and diamond of modal logic. But he also says, more generally, that ''possibility amounts to existential quantification over the worlds''; and ''modality turns into quantification''; and also ''Modality *de re* . . . is quantification over possible individuals''.[16] Here he slides from claims about modal operators (or symbols like boxes and diamonds) to claims about modality itself. I believe that what motivates all of this is the basic idea that certain perplexing linguistic constructions—prominently the modal adverbs—are *eliminable* in favor of simple quantification provided we admit certain sorts of entities into our ontology. But I insist that this could be correct only if the concept expressed by the original construction were none other than the one expressed by the quantificational one. So it would hardly signal the reduction of something modal to something nonmodal.

Returning to the present analytic suggestion, it cannot be overemphasized that Platonic properties are supposed to have fixed intrinsic natures. How they are in themselves depends in no way on the concrete realm. *Being a horse* would be no different intrinsically if there had been no horses, no different if all horses had been wild, no different if some had been blue, and so on. Similarly for *being an animal*. That one property entails another is a matter of fact on all fours with the fact that one horse is taller than another. Of course the former is a fact about the Platonic realm while the latter is not. So, where modal propositions may once have seemed to transcend the *actual*, they now seem only to transcend the *concrete*. The boundaries of the possible are not determined by nonactual circumstances, but rather by actual relations among properties. That horses must be animals does not depend on there being no inanimate horses among ''nonactual entities'', but rather simply on the entailment that we've been discussing.[17]

[16] See section 1.2 of *On the Plurality of Worlds* (Oxford: Blackwell, 1986).

[17] Another important intrinsic relation between properties is the (symmetric) relation of *compatibility*. Properties are compatible, intuitively, if their intrinsic natures allow for something to instantiate both. Thus *being round* and *being green* are compatible. So are *being female* and *being president*; *being Venus* and *being unnamed*, and *being a horse* and *being wild*. Obviously entailment can only hold between compatible properties, but these examples show that compatibility doesn't guarantee entailment. Given abundance, it's easy to see that compatibility and entailment are interanalyzable in the manner of necessity and possibility, so that either notion could be taken as primitive. Compatibility of properties, of course, is already familiar. But, as with entailment, the usual approach to the notion is backwards: we are told that the compatibility of two properties consists in the possibility of their both being instantiated by one thing.

5. ALL THE THINGS YOU ARE

At this point it may seem that although the entailment of properties might offer a promising basis for analyzing *de dicto* modality, the strategy cannot be extended to the *de re*. But I'll now try to show that it can. Along the way I'll urge that our new analytic tools reveal that typical nontrivial claims of *de re* necessity are in fact false, and that it's a fundamental error to identify the doctrine of essentialism with non-trivial necessity *de re*.

Quine argued that although we might be able to make some sense of the modal notions as applied to *sentences*, we can't make sense of modality as applied to *things*. A well known example goes like this. Suppose we agree that it's necessary that cyclists are two-legged and mathematicians are rational. (Try not to let the actual implausibility of these claims distract.) Now suppose we have on hand someone who is both a cyclist and a mathematician. If we describe this person merely as a cyclist, then he will necessarily be two-legged but he won't necessarily be rational. But the opposite holds if we describe him merely as a mathematician. Quine concluded that the attribution of an essential property to a thing made sense only with respect to a prior description of the thing—that the unqualified attribution of an essential property to a thing was nonsense.[18]

I think that this is a very interesting argument. For one thing, Quine's extensionalism aside, it virtually concedes that genuine modality at least sometimes resides in relations between properties. In effect, *being a cyclist* is being imagined to entail *being two-legged* (and *being a mathematician* to entail *being rational*). The cyclist's being necessarily (or essentially) two-legged is taken to consist in this entailment, and I think this is fundamentally correct. But then there's a crucial mistake. For Quine assumes that to attribute (say) rationality to the *thing* independently of its being a cyclist or mathematician (etc.) would be to do something utterly different from attributing the property to it under some such description. Typical critics of Quine don't question this, but they do urge that the argument fails, insisting that the utterly different notion makes perfectly good intuitive sense after all.[19] I think that everyone here is at least partly wrong. The argument

[18] See *Word and Object*, 199.

[19] For example, see Kripke's remarks (using different examples) in *Naming and Necessity*, 39–42.

fails, but the notion is *not* utterly different. In fact, the failure of the argument traces directly to the fact that the notion isn't (relevantly) different at all.

To be a specific physical thing, on the present account, is nothing more nor less than to instantiate a certain object-essence. Setting dualism aside, suppose x is the physical object that is our cycling mathematician. Just as *being a cyclist* entails certain properties and fails to entail certain others, so it is with the object-essence *being x*. For example, as a result of mereological essentialism, if z is a part of x, then *being x* entails *having z as a part*. Of course, *being x* doesn't entail *being rational* or *being two-legged*. But this hardly shows that it *doesn't make sense* to assert that these properties are essential to x. To the contrary, it makes perfectly good sense, as the example of *having z as a part* reveals. The problem is just that it's *false* that x is essentially two-legged or essentially rational. As I will analyze them, these claims are false precisely because these properties are *not* entailed by *being x*.

I am therefore suggesting that the *de re* is to be analyzed in terms of the *de dicto*—that the former is just a special case of the latter.[20] Roughly speaking, we have a case of modality *de re* when (and only when) the appropriate property in the *de dicto* formulation is an *object*-essence. Let's illustrate this using a famous Kripkean thought experiment. Suppose a certain physical object w instantiates a certain table-essence, the property of *being such and such table*, and that w is in fact made of wood. To make the intuitive *de re* claim that w is necessarily (or essentially) made of wood, on the present view, is to say that its object-essence, *being w*, entails *being made of wood*. But (suppressing the existential presupposition) this entailment is, on the present view, precisely the truth-condition for the classic *de dicto* formulation, ''necessarily, for any x, if x instantiates *being w*, then x instantiates *being made of wood*''. And to say that the *table* is necessarily made of wood is to say that the table-essence *being this table* entails *being made of wood*. This, of course is the *de dicto* claim that it's necessary that anything that instantiates *being this table* also instantiates *being made of wood*.

[20] This is not a novel suggestion. For example, Alvin Plantinga offers a ''special case'' treatment in *The Nature of Necessity* (Oxford: Oxford University Press, 1974). The details are rather different, reflecting a different understanding of the nature and role of proper names.

Now, *being w* simply does not entail *being made of wood* since the parts of w could have been arranged so that it wasn't made of wood. So it isn't necessary that anything that instantiates *being w* also instantiates *being made of wood*. But if the familiar Kripkean thought experiment is right, and we couldn't have *this table* without its being made of wood, then the table-essence *being this table* indeed does entail *being made of wood*. Hence, to infer—à la Kripke—that the *object* w is essentially made of wood from the intuited premise that the *table* is essentially made of wood is mistaken. On the present analysis, it would be to infer that *being w* entails *being made of wood* from the premise that *being this table* entails *being made of wood* (given that w instantiates *being this table*).

I believe that a careful consideration of the entire gamut of Kripke's ''essentialist'' thought experiments reveals that while they indeed offer persuasive support for conclusions of the latter sort, they offer none at all for the former. In fact, I think that Kripke drew the proper conclusions from the experiments, but then fell into erroneous *de re* claims by what he took to be innocent paraphrase.[21] What seemed to be innocent was anything but, and it was a dramatic case of object fixation. A very important consequence is that it's a mistake to follow Kripke in thinking of essentialism as the doctrine of nontrivial *de re* necessity.

In Section 1 I urged that the real source of our intuitions about necessity and possibility is our thoughts about relations among properties. I believe that this claim is well supported by a careful consideration of Kripke's essentialist thought experiments. Now is a good moment to explore this further. It's a commonplace among essentialists that *being a person* is essential to each of us. A Kripkean thought experiment that favors this conclusion might run along the following lines. First we ask, say, whether we can imagine ''having Nixon without having a person''. Now, just what do we do in trying to imagine this? Suppose we first imagine a yacht. Perhaps we say to ourselves, ''Well, that's not Nixon''. Then maybe a palm tree, with the same conclusion. An alligator? A zombie? No. Then perhaps we turn things round. We conjure up a

[21] Notice how, on p. 47 of *Naming and Necessity*, he moves from ''this table'' to ''this very object'' without missing a beat in pressing the essentialist claim: ''*This table* is composed of molecules. Might *it* not have been composed of molecules?. . . it's hard to imagine under what circumstances you would have *this very object* and find that it is not composed of molecules.'' (My emphasis.)

visual image of an entity that looks a lot like Nixon. Maybe a life-size cardboard Nixon-prop, or an automaton. Well, again, that's not Nixon. Finally we imagine someone who looks and acts a lot like Nixon, and gives no sign of being otherwise. We've somehow imagined Nixon, and yes, in doing so we've imagined a person. The net effect of our imaginative efforts is that we feel Kripke is right: we can't imagine Nixon failing to be a person.

But exactly why are we entitled to the initial negative conclusions? For example, when we imagine the palm tree, what is it that guarantees we haven't imagined Nixon? I think there is only one plausible answer: At some level, we know that *being a palm tree* is incompatible with *being Nixon*, and the imagining simply helps make this vivid. The same holds for the yacht, alligator, zombie, prop, and automaton, and would hold for any sort of thing that isn't a person. It's important to see that it would be ludicrous to think that our imaginings somehow ensured that the palm tree (etc.) wasn't the *physical object*, say x, that actually happens to instantiate *being Nixon*. Here we have no way of thinking about x in isolation from the property of *being Nixon*.

Finally, why are we entitled to conclude that in the last part of our thinking, we actually *did* "imagine Nixon"? Say we formed a visual image of a man wearing a dark suit, with a certain sort of hairline and a forced smile, about to enter a helicopter, arms raised so as to form a 'V', and so on. Certainly nothing about the visual image *itself* made it an image of Nixon. This very image would have worked fine had we been asked to imagine an actor in a movie doing a scene depicting Nixon's final departure from the White House. Just as clearly, there was nothing in the image itself that made it an image of the physical entity x.

It can only be that conjuring up this image counts as imagining Nixon—when it does!—because part of the conjuring is our intention that it represent Nixon. (Certainly not the intention that it represent x.) The property of *being Nixon* is therefore intimately involved in our being able to imagine Nixon. We cannot give our imaginings the correct representational force unless this property plays a role in our intentions.

The essentialist thought experiments are therefore nothing more than picturesque reflections on the entailments and compatibilities of properties. They are (generally) *not* thought experiments about specific physical objects. They help make it clear that *being Nixon* entails *being a person*, is incompatible with *being a palm tree*, and so forth. There is nothing at all in these experiments to warrant any similar conclusions

about the physical entity x. As far as the experiments go, *being x* may very well be compatible with *being a palm tree*, and it certainly doesn't entail *being a person*.

As suggested earlier, some philosophers may not be entirely enthusiastic about the present view of *stuff* as a primitive concept. I promised to mention an alternative view, and this is a good moment to do so. It's an ''atomistic'' view. All matter is conceived to be composed of ''fundamental particles'' (or packets (or very local fields) of energy, or modifications of spacetime, etc.). There is a (possibly infinite) list of varieties of fundamental particles (or fields, modifications, etc.). These are just properties whose instances, if any, are best thought of as tiny regions of spacetime. If the ''fundamental particles'' really are just ''modifications''—attributes of spacetime regions—then these attributes are on the list. If, on the other hand, it really makes sense to speak of particles (or fields, etc.) as distinct from the regions they occupy, then the relevant properties are (speaking loosely) of the form ''being occupied by such and such sort of particle''. The list, of course, may include properties of both sorts, and may include properties that are not exemplified in the actual world, given the way matter actually happens to work.

Now, the present approach to modality relies on there being, for any physical object at all, a property of being that specific physical object. I called such properties object-essences. What can such properties be like when the objects in question are fundamental particles (or modifications, etc.)? I am going to assume, I think plausibly, that to be a fundamental particle (etc.) of any specific sort (i.e. to instantiate some property on the list) is a strictly *intrinsic* affair: that such a property is instantiated by a particle or region strictly in virtue of how that particle or region is in itself. I also assume, again I think reasonably, that no two particles (etc.) of the same fundamental variety have any intrinsic differences. Thus the object-essence for such a particle is a complex property that is partly intrinsic—reflecting its fundamental type—but also partly relational. I am going to take the relational component to be strictly spatiotemporal. If we conceive of fundamental particles as very thin spacetime worms, then a given such worm either has its beginning at some small spacetime region or else is backwards-eternal.[22]

[22] Here I will be adopting a classical perspective. Adapting it so as to conform with typical understandings of quantum theory would require probabilistic complications.

In the former case, let the given particle have spacetime origin R. Then we may take its object-essence to be the property of *being a particle of such and such type originating at spacetime region R*. This decision makes the origins of particles essential to them, but allows that they might have subsequent spacetime histories that differ from their actual histories. I believe this is intuitively correct. It is an ''anti-haecceitist'' position regarding fundamental particles. Two fundamental particles of the same type could *not* have had their spacetime origins interchanged.

We deal with the backwards-eternal case as follows. First, for any given type T of fundamental particle, there is a definite class of regions of spacetime that are suitable for (precisely) containing a particle of type T. We call such regions T-paths. (Most of them are of course not occupied by T-particles.) Now consider the relation that one T-path bears to another if there is some time prior to which they entirely coincide. Such paths are called *origin-indistinguishable*.[23] Obviously origin-indistinguishability is an equivalence relation, so it partitions the class of all T-paths into equivalence classes. Each equivalence class is, in effect, the set of all possible spacetime histories for a specific particle of type T. Thus we may think of each equivalence class as a specific *origin-property* for a T-particle. If we have a given T-particle, its object-essence will then be the property of *being of type T and having such and such origin-property*. The second conjunct is, in effect, the property of *having a path belonging to such and such equivalence class*. So a given T-particle could have had very different spacetime histories, but could not have had backwards-eternal differences from its actual history. This approach therefore adapts the doctrine of the essentiality of origins to the backwards-eternal case.

A welcome consequence of this treatment is that we easily obtain *non-mysterious* uninstantiated object-essences for fundamental particles. For example, in the second case, there was no requirement that every equivalence class include a path that is occupied by a fundamental particle. But the property of *being of type T and having a path belonging to such and such equivalence class* is still a perfectly good property whether instantiated or not. If it is instantiated, it's the object-essence of a fundamental particle. If not, it remains non-mysterious because

[23] Paths are *origin-distinguishable* if for any time t, there is a prior time t* such that the t*-slices of the paths are distinct.

it isn't a (non-qualitative) haecceity. (For the easier first case we simply consider type-appropriate originating regions of spacetime.)

Now, from this atomistic perspective, any *non-fundamental* object is just a mereological sum of parts of fundamental particles that happen to be assembled in a certain way. Thus we may take an arbitrary non-fundamental object-essence to be the property of *being the mereological sum of such and such parts of such and such fundamental particles*. Generalizing to uninstantiated non-fundamental object-essences is straightforward.[24]

I believe that the metaphysical results of adopting this picture are the same as those of the ''primitive stuff'' picture. In fact, I think of the primitive (-but-not-*simple*) stuff picture as a generic picture that accommodates the present picture and others as well. Under the atomistic picture it's still true that different objects could have instantiated such properties as *being Nixon*, and hence that typical examples of alleged *de re* modality are unpersuasive. Finally, mereological essentialism remains unscathed by standard examples based on objects of familiar kinds.

6. AT LAST

We have now seen how entailment between properties is held to provide the analysis of typical cases of modality, whether *de dicto* or *de re*. I believe that the entailment approach can be systematized in a more general and formal way, but doing so is beyond the scope of this chapter. The usual procedure would be to introduce a formal language, to show how to interpret it in a mathematically precise way, and then to argue that ordinary modal claims may be adequately translated into the interpreted system. It is not difficult—only tedious—to do this for the entailment analysis. But any such formal approach to modality is unlikely to be entirely satisfactory. The reason, as David Lewis nicely put it, is that ''modality is not all diamonds and boxes''.[25] There are

[24] What is required is simply that the object-essence involve at least one uninstantiated fundamental-particle object-essence. It is worth noting that having non-haecceitistic (and hence non-mysterious) object-essences is a major difference between the current metaphysical view and the one defended by Plantinga in *The Nature of Necessity*.

[25] *On the Plurality of Worlds*, 13.

everyday modal idioms, and also everyday modal truths, that do not translate easily (or even at all) into the language of quantified modal logic, or are not fully (or in some cases even remotely) captured by typical semantical treatments of such a language. Examples mentioned by Lewis include modalized comparatives, supervenience, and, of course, counterfactuality. He also shows that although his own formal account of modality can be used to reflect most of our intuitions about essential properties in a systematic way, the very same devices that achieve this outcome give the wrong results if applied in the same way to the curious case of necessary existence. His moral, in effect, is that we should be willing, case by case, to step outside the formal treatment and just exploit the fundamental analytic ideas directly to say what is going on in exotic, recalcitrant, or otherwise special cases. I believe Lewis was correct in these observations, and that they apply not just to his own treatment, but to any system whose central project is analyzing necessity and possibility. That, of course, would include the present approach. So in this final section, rather than detailing a formal system, I'll simply say a little about what its language would be like and then consider how a few examples would or would not be well accommodated.

In discussing the great divide and object fixation above, I claimed that certain modal claims are not just ''about'' physical objects, but also ''about'' *properties*. I also emphasized the Platonic principle that (e.g.) for something to be a horse *just is* for it to instantiate the property of *being a horse*. The natural way to honor this principle is to understand even simple nonmodal sentences as expressing (intuitive) propositions that have properties as constituents. To say that something is a horse is accordingly to say that a certain relation holds between it and the property. So the *property* is just as much a constituent of the proposition as is the *thing* (if, indeed, the thing proves to be a constituent).

Thus consider a sentence like 'Alf is a dog'. (Imagine an actual dog.) It would be represented in first-order logic by an atomic, predicate-letter/nonlogical-constant formula, say 'Da'. But from the present perspective this is entirely inadequate. There is a fatal problem with the 'D' and also a fatal problem with the 'a'. Translating 'is a dog' by the predicate letter 'D' threatens to miss the constituency of the property because the predicate position is immune to quantification. We need to be able to infer 'There is something that Alf instantiates' from 'Alf is a dog' since, on our Platonic principles, this sentence just means that Alf bears the instantiation relation to (the property) *being a dog*.

Now for the 'a': I argued in Section 3 that ordinary proper names like 'Alf' aren't ''singular terms'' in the usual sense, notably are not ''rigid designators'', and that instead they express properties. I claimed (waiving defense) that the properties they express are k-essences. So it will not do to translate 'Alf' by an individual constant like 'a'.[26] 'Alf' expresses a specific dog-essence. Assuming a context in which 'Alf' has existential import, this view of names could be accommodated by adopting the translation '∃x(Ax & Dx)', where 'A' expressed (the dog-essence) *being Alf*, but this would still leave the 'D'-problem unaddressed.

We have concluded that when we say that something is a dog, we are covertly referring to the property of *being a dog* in a way that supports quantification, not just ''expressing'' it in some weaker sense. One way to allow for this is to move to second-order logic, allowing quantification over the predicate position. But I prefer to stick with first-order logic by making the instantiation relation explicit and translating English predicates and proper names by individual constants for properties.[27] Thus a name is translated by a constant for the k-essence of its bearer, not by a constant for the bearer. Using 'I' for instantiation, 'd' for *being a dog*, and now 'a' for *being Alf*, we arrive at the translation '∃x(Ixa & Ixd)'. So 'Alf is a dog' does not express a ''singular'' proposition (with respect to 'Alf'). The specific entity that happens to instantiate *being Alf* is thus not a constituent of the expressed proposition. (*Being Alf* and *being a dog*, of course, are constituents of the intuitive proposition.) So the first step toward systematization (in the present first-order style) would be to modify standard modal languages by dispensing with unary predicates in favor of a binary relation for instantiation and (''directly referential'') individual constants for properties.

With this mild reform in mind, I would now like to show how it may be exploited to capture a species of modal claim that is distinct from though typically confused with modality *de re*. In my view such claims are supported by ''essentialist'' intuitions that in fact do not support nontrivial modality *de re* (as detailed earlier).

Consider the sentence 'Dogs are necessarily canine'. It is surely ambiguous in everyday English. I believe it has at least four readings,

[26] I would not deny that we are capable of introducing proper names that ''refer directly'' to specific entities, only that this is not the way ordinary proper names function.

[27] The first-order treatment makes it more vivid that properties and physical objects enjoy the same status as regards intuitive propositional constituency.

and that we can gain a better sense of the power of the entailment-based analysis by examining three of them.

The familiar *de dicto* reading of this sentence is more strongly suggested by 'It is necessary that all dogs are canine'. We may translate this into the modified formal language by the formula

(1) $\Box \forall x(Ixd \rightarrow Ixc)$,

with 'd' for *being a dog* and 'c' for *being canine*. Intuitively, (1) says something that ought to be true on any reasonable interpretation of the box. On the analysis, the truth condition for (1) is simply that *being a dog* entails *being canine*. Typical English sentences that genuinely merit the *de dicto* treatment would have translations akin to the form (1).

Next we have the parallel and equally familiar *de re* reading, and we formalize it in the expected way:

(2) $\forall x(Ixd \rightarrow \Box Ixc)$.

According to the entailment analysis, (2) will be true iff each thing that instantiates *being a dog* is such that *its object-essence* entails *being canine*. It should be clear from what has already been said that I believe this *de re* reading is false, and in fact that *no* thing that happens to be a dog is necessarily canine (nor necessarily an animal, etc.). In other words, I believe that no object that happens to instantiate *being a dog* is such that its object-essence entails *being canine*. Any English sentence that *genuinely* deserves the *de re* treatment is treated along the lines of (2).

But there is a third reading. As suggested above, I think it is supported by a careful understanding of the thought experiments. For one way to understand 'Dogs are necessarily canine' is to take it as a certain kind of generalization from sentences like

(3) Alf is necessarily canine.

What I have in mind is a generalization from a certain *kind* of thing, of which Alf is an instance, to all cases *of that kind*. Since Alf is a dog, *being Alf* is a dog-essence and the generalization ''from Alf'' is therefore—in a way to be described—a generalization to all dog-essences.

Imagine a Kripke-style thought experiment about the dog Alf: Can we conceive of having Alf without having a canine? I urged above that what is really going on in such a thought experiment is a comparison of properties. We are asking whether *being Alf* is compatible with *not*

being canine (or, in other words, whether *being Alf* entails *being canine*). If we decide that indeed we cannot imagine Alf failing to be canine, it can only be because we take *being Alf* to entail *being canine*. *It is this entailment that supports our feeling that (3) is true*. So now let's work backwards to figure out how (3)—given that it's true—is best translated into the modified formal language.

On the face of it, there are four potential translations. The first two are:

(4) $\Box\exists x(Ixa \,\&\, Ixc)$;

and

(5) $\exists x(Ixa \,\&\, \Box Ixc)$.

But neither of these can be right under the present principles. (4) is false because Alf doesn't exist necessarily. (Below we will see how to analyze necessary existence.) And (5) is false because there is no *thing* whose object-essence entails *being canine*. The remaining two possibilities differ only as to existential import:

(6) $\exists x(Ixa \,\&\, \Box\forall y(Iya \rightarrow Iyc))$;

and

(7) $\Box\forall y(Iya \rightarrow Iyc)$.

As we have imagined the example, Alf is a specific actual dog. So *typical* uses of 'Alf' carry existential import and the correct translation of a typical use of (3) is therefore (6). If Alf had been a fictional dog, the import-free (7) would have been more appropriate.

Being Alf, of course, entails *being canine* whether Alf happens to exist or not. Thus, although an ordinary utterance of (3) is best captured by (6), the content of the thought experiment *itself* doesn't depend at all on Alf's actually existing. We could just as easily do it with *Goofy*. It therefore seems that if we are contemplating a generalization of sentences *generally* like (3), we do better to take it as translated by (7). To put it a little differently, in order to obtain full generality it is better to imagine that *being Alf* is an arbitrary dog-essence rather than one we know to be instantiated.

Now let's assume that sentences like (3), *to the extent that they reflect essentialist thought experiments*, are best translated into the modified language by formulas like (7). And let's agree that there is a sense of

(8) Dogs are necessarily canine

in which it is a kind of generalization of (3). (A helpful way to think of it is suggested by: ''Any specific dog is necessarily canine.'') Intuitively speaking, (7) reflects the thought experiment by taking *being canine* to be essential to ''Alf himself''. This is mirrored directly in the analysis, for (7) will be true iff *being Alf* entails *being canine*. It *isn't* that *being a dog* entails *being canine*, even though that of course is also true. Now, there is surely a sense of (8)—in fact the *de dicto* sense given by (1)—that is appropriately captured by this latter entailment. But this isn't the sense we seek. We want the sense in which (8) says that *being canine* is essential to ''any dog itself'', just as (3) says that *being canine* is essential to ''Alf himself''. (This sense of (8) may be more forcefully suggested by 'Any dog is itself essentially canine'.) We therefore generalize to all dog-essences. One way to do this is to introduce a function symbol, say '*', expressing a function that maps an arbitrary kind, k, to the property of *being a k-essence*. Then we may translate (8) by

(9) $\forall x(Ixd^{*} \rightarrow \Box \forall y(Iyx \rightarrow Iyc))$.

On the entailment analysis, (9) is true iff every dog-essence entails *being canine* (again with no restriction to instantiated dog-essences). I call the reading given by (9) the *essentialist* reading.

We have now seen that modal sentences like (8) have three fundamental interpretations: *de dicto*, *de re*, and *essentialist* (along with variations depending on whether the context imposes existential presuppositions). An obvious moral is that English sentences do not always wear their proper formal translations on their faces. Translation is always a matter of capturing an intuitive proposition, and there is no routine and mechanical way to achieve this. We simply have to think about each case, and do what seems best.

Now let's turn to the important example that forced Lewis to depart from his straightforward systematic deployment of possible worlds and counterparts as analyzing the boxes and diamonds: the strange case of *necessary existence*.

Thus consider the proposition that some (given) entity, say the recent Alf, exists necessarily. The natural translation is '$\exists x(Ixa \;\&\; \Box Ixe)$', with 'a' again for *being Alf* and 'e' for the property of *existing*. It will not do to treat this case uniformly with typical *de re* cases by taking its truth condition to be the entailment of e by the object-essence of the thing that in fact instantiates a. The reason is that, on intuitive actualistic and Platonic grounds, *every* property entails *existing* (including

uninstantiated properties like *being a unicorn* or even *being a round square*). This principle is no less appealing for object-essences than for familiar general properties, for it merely reflects the very natural idea that only things that exist can be instances of properties. The challenge is thus to provide a plausible account that exploits the entailment idea even as it departs from the systematic treatment that applies successfully to overtly similar cases.

For Lewis the problem was essentially the same. His overall treatment of intuitive necessary properties was that a thing had a property necessarily iff every counterpart of it had that property in its own possible world. (That the entity-essence of the thing entails the property is the present analogue.) But then along came Humphrey and his counterparts, and of course every counterpart of Humphrey exists in its own world. (*De re* analogue: every entity-essence entails existing. *Essentialist* analogue: every person-essence entails existing.) Yet neither Humphrey (nor any physical object) exists necessarily. Lewis responded by applying the ingredients of the counterpart apparatus informally (and non-uniformly) in the special case of necessary existence: to exist necessarily was to have a counterpart in every possible world. This was certainly the right thing to say from his perspective.

Many philosophers have thought that some things ''exist necessarily'' (or at least that it's an epistemic possibility that some things exist necessarily). For example, I think there really are necessary entities, specifically the Platonic properties and relations, and also that no concrete entities exist necessarily. Although I have reasons for this opinion, I won't offer them nor rely on the opinion here. What is required for an analysis of necessary existence is a perfectly general property that some entities may or may not in fact have, and which plausibly reflects our intuitive concept of necessary existence. But here it is very important to notice that this intuitive concept is in fact extremely vague. We don't really have an intuitive account of it that goes beyond restatement. (It just doesn't help to say that for something to exist necessarily is for it to be such that it could not have failed to exist. Or to say that it exists in all possible worlds, or that it exists no matter what, etc.) But this vagueness may actually make the task somewhat easier, for success then seems to depend more on overall fit with the analytic framework than on matching some reasonably clear pre-existing notion.

It's also very important to appreciate that in seeking an analyzing property we are *not* seeking a property that somehow guarantees its

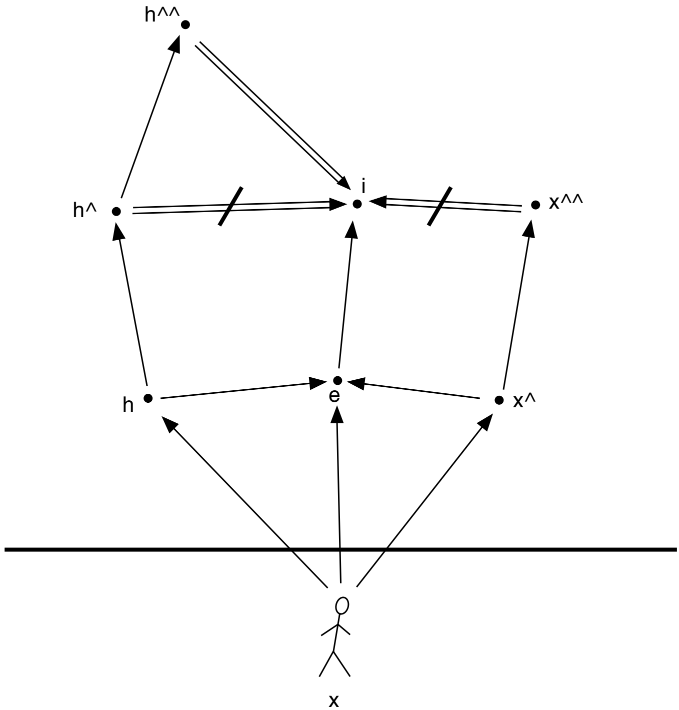
own instantiation, that guarantees that something *else* exists just by existing itself. Our problem is rather this. We are already *given* what exists. The question is whether, among the things that *do* exist, there is some feature that (perhaps) some of them have and others don't have, a feature that ''governs'' the necessity of their (undisputed) existence. We have just seen that *having an entity-essence that entails existing* does not achieve this, roughly because it would follow that everything exists of necessity.

I believe that the correct analysis is this: for a thing to exist *necessarily* is for it to have an entity-essence whose entity-essence entails *being instantiated*. Using '$x^{\wedge}$' to denote the entity-essence of x (for any x), and 'i' for the property of *being instantiated*, we may capture this by saying that x exists necessarily iff $x^{\wedge\wedge}$ entails i. It is crucial to see that it doesn't follow from the fact that $x^{\wedge}$ entails *existing* that $x^{\wedge\wedge}$ entails i. Thus we leave room for the (epistemic) possibility that some entities x are such that $x^{\wedge\wedge}$ entails i but others are not, and this despite the fact that y entails *existing* for any entity-essence y (whether instantiated or not). Thus, if the opinion expressed above is correct, then no concrete entity x is such that $x^{\wedge\wedge}$ entails i, but every Platonic entity x is such that $x^{\wedge\wedge}$ entails i. Figure 4.1, with concrete entities below the line and Platonic entities above, provides a pictorial representation of this opinion by way of the Humphrey example, thereby illustrating the main idea. In the diagram single arrows represent instantiation, double arrows entailment, and 'h' denotes the person-essence, *being Humphrey*. (The diagram reflects the necessity of the property h's existence as well as the contingency of Humphrey's existence and x's existence.)

I've tried to present this proposal in a way that might forestall (or disarm in advance) a certain skeptical reaction, but I may not have succeeded. The reaction is that $x^{\wedge\wedge}$'s entailing i is just the holding of a certain intrinsic relation between a pair of properties, and that surely no such state of affairs could guarantee that anything exists (except perhaps the properties themselves), so that it cannot be that $x^{\wedge\wedge}$'s entailing i is what it is for x to exist necessarily.[28]

I believe that this reaction is undermined by a pair of closely related problems. First, notice that it presupposes that we already have a reasonably clear intuitive concept of necessary existence, for it claims

[28] I thank Kirk Ludwig and Greg Ray for forcing me to deal with this question.



that the proposal falls short of capturing it. But I have claimed that we in fact have no such clear pre-analytic concept. I am offering a proposal to answer to what is in fact a vague and misleadingly analogical concept—the result of imagining that 'necessarily', as it modifies 'existing', somehow expresses the same basic modal notion it expresses when it modifies a more humdrum adjective like 'human'. But the concept of *existence* is already an essential component of our conception of intuitive modal properties like *being necessarily human*. This should alert us to the possibility that exactly the *same* modal qualification may not be readily applicable to the concept of existence itself, and that any felt analogy between *necessary humanness* (etc.) and *necessary existence* may just be a misleading byproduct of an accidental grammatical parallel. If something could not have existed without being human (etc.), that is far from trivial, but it's entirely trivial that nothing could have existed without existing. This suggests very strongly that *necessary humanness* and *necessary existence* are not two manifestations of a

single modal phenomenon, and hence that it's no surprise if they require different analyses. So the first problem with the reaction is that it takes it that necessary existence is sufficiently clear pre-analytically for us to see that the proposal falls short of capturing it.

The second problem is this. The objection gets its rhetorical bite from the suggestion that whereas relations between properties are just inert Platonic states of affairs, the concept of necessary existence has very serious ontological heft. Thus it is made to seem absurdly magical that necessary existence should be grounded in static eternal relations between *abstracta*. Now of course it is right that nothing springs into existence as a result of intrinsic relations holding between Platonic properties. But what we're trying to capture here is the *necessity* of the existence, not the *fact* of the existence. If there are any necessary beings, then they are no different from contingent ones as far as their mere *existence* goes. For a necessary entity to exist is nothing more nor less than what it is for a contingent entity to exist.[29] Here we are discussing things whose existence is already given, and we're asking whether, among them, there is some *further* feature that (at least potentially) sets some apart from others in a way that reflects our vague intuitive idea that their existence is somehow *necessary*. So the ontological heft is already there, presupposed in our inquiry, and we're looking for something over and above that ontological heft. I believe that when we see the matter from this perspective the objection loses its rhetorical force. The present suggestion is then that there is in fact a certain *property* that sets the necessary existents (if any) apart from the contingent ones (if any!), namely the property of *having an entity-essence whose entity-essence entails being instantiated*.

This approach to necessary existence has what I think is a welcome companion application in the metaphysics of fiction. I have defended the Kripkean view that the intuitive ''worlds'' of works of fiction are not metaphysically possible.[30] (Here it is best to think of the world of a fiction as a complicated proposition that the fiction *expresses* but doesn't *assert*.) Thus although it is possible that any logically consistent story be ''enacted'' (and possible for it to be enacted by many different ''casts

[29] To emphasize: to *say* of a necessary entity that it exists is to say the very same thing of it that one says of a contingent entity in saying that *it* exists.

[30] See chapter 10 (''Is There Truth in Fiction?'') of *Contemporary Metaphysics* (Oxford: Blackwell, 1997).

of characters'') it is not true that the fictional names in the story denote any of the entities involved in any of the (intuitive) possible enactments (notably not in the case when it's actually enacted). On this view, a fictional name like 'Sherlock Holmes' does not express a person-essence that could have been instantiated. Yet in accordance with our overall Platonic picture, we still want *being Sherlock Holmes* (call it 's') to entail *existing* just as any more routine property entails *existing*. The present suggestion is then that we may happily allow s to entail *existing* while at the same time maintaining that Holmes could not have existed, simply by holding that the entity-essence of this person-essence (namely $s^{\wedge}$) entails *being uninstantiated*. The failure of Sherlock Holmes to exist is thus seen as necessary. I believe that even if one does not accept the present view about fiction, reflection on this general approach to necessary *nonexistence* helps vindicate the analysis of necessary *existence*.[31] The reason is that we don't have the distraction of an actually existing entity to mislead us. There is something about the property s that guarantees that it isn't instantiated. It's that it instantiates $s^{\wedge}$, and $s^{\wedge}$ entails *being uninstantiated*. If this is a plausible account of necessary nonexistence, then the structurally indistinguishable account of necessary existence should seem so as well.[32]

[31] For example, those who accept ''inconsistent'' properties, like *being a round square*, may hold that they entail *existing*, but that their entity-essences entail *being uninstantiated*.

[32] Material in this chapter was presented at the seventh Metaphysical Mayhem conference at Syracuse University in August 2002, at the University of Florida in March 2003, at a session at the APA-Pacific meetings in Pasadena in March 2004, and at the 56th Northwest Philosophy Conference in Bellevue, Washington in October 2004. I am very grateful to Greg Fitch, who commented at Syracuse, to Karen Bennett and Linda Wetzel, who commented in Pasadena, to members of the four audiences, and also to colleagues and students, for helpful comments. These many people include Jim Blackmon, Gina Calderone, David Copp, Greg Damico, Andre Gallois, Dan Gaskill, Geoff Georgi, Peter Hanks, Thomas Hofweber, Jeffrey King, Kirk Ludwig, Frank McGuinness, Trenton Merricks, Greg Ray, Steve Reber, Ted Sider, Paul Teller, and Dean Zimmerman.

This page intentionally left blank

Part III

COINCIDENT OBJECTS AND TEMPORAL PARTS

This page intentionally left blank

5. Four-Dimensionalism and the Puzzles of Coincidence

Matthew McGrath

1. INTRODUCTION

Often cited in defense of four-dimensionalism about the persistence of material objects is its treatment of the so-called *puzzles of coincidence*. These puzzles include the statue/lump, the ship of Theseus, Tibbles the cat, and the various fission and fusion puzzles in the personal identity literature. In their original versions, the puzzles involve changes which seem either to produce or terminate coincidence between material objects (the lump is flattened, the cat's tail is cut off, etc.), but each of the puzzles also has a modal variant in which the relevant change could have occurred but does not. Four-dimensionalists (4Dists) standardly take themselves to have an edge over three-dimensionalists (3Dists) in the treatment of these puzzles. They claim that the original puzzles are answered easily and painlessly under 4Dism, and that their modal variants can be answered by something like counterpart theory. By contrast, they claim, 3Dists have no easy way with the originals, and no better way with the modal variants.[1]

My aim here is to determine whether this is correct. I will argue that it is not, and in fact that the puzzles are every bit as challenging for the 4Dist as they are for the 3Dist. In a final section, I will tentatively suggest that reflecting on the puzzles may provide us with reason to reject 4Dism.

[1] This view is widespread among 4Dists and 3Dists alike. See, e.g. Burke (1994), Hawley (2001), Heller (1990, 2000), Lewis (1976, 1986), Noonan (1989), Rea (1997a) (though for a more skeptical take, see Rea 1997b, section 4), and Sider (1996, 1997, 2001, forthcoming).

2. SOME PRELIMINARIES

Before we turn to the main topic of the chapter, we need to characterize 3Dism and 4Dism. This turns out to be more difficult than one would expect. The problem is that two distinct doctrines have been associated with '4Dism' in the recent literature on persistence. The *temporal extent doctrine* holds that material objects[2] persist and that they do so by extending through spacetime along its temporal dimension, or, more picturesquely, by being spatiotemporal worms.[3] The *perdurance doctrine* holds that objects persist and do so by *perduring*, that is, by having temporal parts. As Josh Parsons (2000) has noted, there is room in logical space for the possibility of an object's having temporal extent while not perduring (even if not for the converse possibility). While the temporal extent doctrine does seem essential to any theory deserving the label '4Dism', it is less clear that the same is true of the perdurance doctrine, especially given the fact that it isn't entailed by the temporal extent doctrine.

Similarly, '3Dism' is regularly associated with both the *endurance doctrine* (i.e. objects persist but do not perdure) and the *''no temporal extent'' doctrine* (i.e. objects persist but are not temporally extended). If it is logically possible for an object to have temporal extent without perduring, then it is logically possible for there to be an enduring object that has temporal extent. Must the 3Dist deny temporal extent?

It is worth noting here that an object's having temporal extent is equivalent to its perduring, *modulo* what we may call the *mereological account* of an object's being extended along a given dimension. According to this account, an object is extended along a given dimension iff the

[2] Throughout this chapter, we focus on the persistence of *material objects*. Excluded from the category of *object*, I assume, are at least the following: properties and relations (whether immanent or transcendent), *matter* or *stuff*, *quantities* in the sense of Cartwright (1970), points/regions of space or spacetime, complexes built up non-mereologically from objects, e.g. *sets*, *events* (at least Kim-events) and *facts*, and entities dependent on objects, e.g. boundaries, tropes (on some conceptions), holes. One would like to say that to be an object is to be a substance. A *material* object, I assume, is an object that has a sufficiently large range of material properties, where these include not only the familiar material properties of common sense (size, shape, color, texture), but those of our best science. I will use 'object' throughout to refer only to material objects.

[3] In the context of special relativity, being a spatiotemporal worm is not equivalent to having temporal extent. The latter is frame-relative while the former is not. In fact, in that context, all statements about temporal extent and perdurance require relativization to a frame.

object is located exactly at a region that is extended along the dimension and, for each cross-sectional subregion of that region (with respect to the dimension), the object has a part that is located exactly at that subregion. Given this account, an object is *temporally* extended just in case it is located exactly at a temporally extended spacetime region and it has parts located exactly at each of that region's temporal cross-sectional subregions. This is just a spruced up statement of what it is to perdure. It ensures that an object has temporal parts for each time period of an object's existence.

Given the mereological account, being temporally extended is analogous to being extended along a spatial dimension at a time. I have length, breadth, and height because I occupy a spatial region with these features and have spatial parts at the appropriate cross-sectional subregions. Perdurantists often think of this analogy between temporal and spatial extent as one of the great strengths of their view.[4]

Parsons (2000) recommends that the 4Dist abandon the mereological account, and so face questions such as these: Does having temporal extent require having at least *some* proper temporal parts? Or is being located exactly at a temporally extended region sufficient by itself? If the answer to the latter is affirmative, as Parsons argues, then there can be temporally extended simples. This would provide one way to explain what it could mean to say, as many endurantists do, that a persisting object is *wholly present* at each moment of its existence: all its parts *simpliciter* exist just when it does (because it has only one part *simpliciter*, itself!)[5]

[4] A little more formally:

O is extended along D iff: O is located exactly at some region R such that (i) R has proper D-cross-sectional subregions and (ii) for each D-cross-sectional subregion S of R, O has a part that is located exactly at S.

(We neglect relativity to a given space to reduce clutter.) There is a corresponding mereological account of an object's being extended in a given space.

O is extended iff: O is located exactly at a region R such that (i) R has proper subregions and (ii) for every subregion S of R, O has a part that is located exactly at S.

Given the latter account, extension in 3D space (at a time) is analogous to extension in 4D space: both involve having parts at each subregion of an extended region. The mereological account of spatial extension at a time has come under fire notably from van Inwagen (1981, 1987) and Markosian (1998).

[5] On the other hand, if an endurantist accepts the mereological account of temporal extension, she might well find the ''wholly present'' formulation of endurantism unsatisfactory. See extended note 1 for further explanation.

We will not look further into these questions, except to note that perdurance is not mere ontological baggage for a 4Dist. Without perdurance, the 4Dist is deprived of the apparently easy and painless ''partial overlap'' treatment of the puzzles of coincidence, which is the topic of this chapter. Only with perdurance is the 4Dist guaranteed that the statue and the lump have temporal parts that can be shared. Temporal extent is not sufficient. It is thus no accident that most self-described 4Dists are perdurantists (or at least *minimal-perdurantists*[6]) and that most of them accept the mereological account of temporal extension. In this chapter, we will give perdurantists ownership of the label '4Dism'. A 3Dist, for us, will then be someone who not only accepts endurance, but who denies temporal extent.

Two further preliminary points. First, given what we have just said about 3Dism, the 3Dist need not deny the existence of *stages*, where a stage is an object that exists at and only at a single moment.[7] What the 3Dist must deny, however, is the existence of a *plenum* of stages; she must deny, that is to say, that, for every persisting object and every time at which it exists, it is co-located with a stage (i.e. its spatial region at the time is the same as that of some stage). If she accepts a plenum of stages, she will be hard-pressed to deny that the persisting object at each time has all the same parts as its then-co-located stage. Such part-sharing

[6] One might argue that the (full) perdurance doctrine is not required for the ''partial overlap'' treatment of the puzzles, but only the doctrine of *minimal perdurance*: objects persist and do so in virtue of minimally perduring, that is, in virtue of having moment-sized temporal parts for each moment of their existence. The doctrine of minimal perdurance is equivalent to the doctrine of temporal extension *modulo* a weaker mereological account of temporal extension requiring only that temporally extended objects have parts located exactly at their moment-sized temporal restrictions.

Minimal perdurance without full perdurance, however, is in some ways inconvenient. Many 4Dists would like to say that the statue is a proper temporal part of the lump, for example. But this claim will require separate argument, given only minimal perdurance. Also, many 4Dists accept the fusion axiom of absolute mereology (i.e. any set has a fusion *simpliciter*). This axiom, given minimal perdurance, will have the consequence that all temporal restrictions of an object's region will contain temporal parts. Minimal perdurance may give way to full perdurance.

[7] Some might find it peculiar to require that stages be objects (and so, by our usage, material objects). The essential contrast, though, is between a stage conceived of as a non-mereological (e.g. set-theoretical) complex built from a persisting object and a time, deriving its material properties (to the extent that it has them) from its components and their mode of composition, and, on the other hand, a stage conceived as something more closely resembling persisting objects as standardly conceived. The latter conception, which I rely on in the text, should be amenable to 4Dists. See the postscript to ''Survival and Identity'' in Lewis (1983).

would seem to give rise to a remarkable correlation of properties: I, a persisting person, am seated at t, and so is my co-located stage; I am slumped over, and so is my stage, and so on. What could explain the correlation, if not mereologically grounded property-borrowing of the distinctively 4D kind? What could explain the fact that both I and my co-located stage are seated at t, if not this: I am seated at t in virtue of the fact that I have as a part a t-stage that is seated *simpliciter*?[8]

Secondly, note that figuring in our accounts of 3- and 4Dism is the requirement that objects persist. We have not interpreted this as a claim that *all* objects persist, because we are allowing both sides to be in a position to admit the existence of at least some stages. But if the requirement is merely that *some* objects persist, then our taxonomy will count 3- and 4Dism as both consistent with the denial of the persistence of *ordinary* objects, that is, of the objects spoken of in ordinary life and the sciences, (e.g. statues, lumps of clay, persons, cats, cells, molecules, and atoms (in the physical sense), etc.) Stage theorists such as Sider (1996, 1997, 2001) and Hawley (2001), who accept fusions of stages but deny that ordinary objects are such fusions, would count as 4Dists. Perhaps this is as it should be. We will skirt this particular taxonomic question, and say only that our interest lies with versions of 4Dism (3Dism) according to which ordinary objects persist, and the persistence of any object, ordinary or not, is a matter of perdurance (endurance).

3. THE PUZZLES OF COINCIDENCE

With these characterizations of 3- and 4Dism in place, we turn to the puzzles.

I begin with some terminology. Let us say that x and y coincide at t iff x and y are distinct but share all parts at t, and that x and y diverge at t

8 Those familiar with the exchange between Sider (2003) and Koslicki (2003) will notice that I am siding with Sider on the issue of whether the acceptance of universalism about diachronic composition commits one to the acceptance of 4Dism. One might reject Sider's definition (2001: 59) of instantaneous temporal parts—''x is an instantaneous temporal part of y iff there is a t such that x exists only at t and is a part of y, and for every object z existing only at t, z is a part of y only if z is a part of x''—on the grounds that satisfying the *definiens* does not, by itself, entail that the persisting object has the t-stage as a part *simpliciter*. But if all objects have Sider-temporal-parts at each moment of their existence, it is very difficult, for reasons we have explained in the body of the chapter, to resist the 4Dist conclusion that they have the relevant stages as parts *simpliciter*.

iff they either become coincident at t or cease to be coincident at t.[9] Past-to-future divergence is fission, future-to-past is fusion. We may classify the original puzzles, then, as either fission or fusion puzzles, depending on whether they threaten fission or fusion. Because our concern in this chapter is with the original temporal puzzles rather than their modal variants, we will mostly neglect modal matters.

It will help to have a representative pair of cases. I choose these, partly for their structural similarity:

> *Statue/Lump Fission Case*: Before us is a statue of Goliath made of soft clay, and so a statue and a lump of clay. Next, we apply a flattening blow. It would appear that we have destroyed the statue but not the lump.

> *Statue/Lump Fusion Case*: Before us is a shapeless lump of clay. We mould it carefully to take a certain shape, resembling Goliath. It would appear that we have created a statue but not a lump.

Reflecting on these cases, we have little trouble constructing strong arguments for the claim that they involve coincidence. But we are also puzzled by coincidence, and with some effort, we can justify or at least explain our puzzlement by constructing arguments. We will briefly discuss pro-coincidence arguments for the two cases above, and then consider at considerable length one of the best anti-coincidence arguments. Our reason for devoting significant space to the latter is that we want to know why coincidence is so problematic, and how, if at all, 4Dists can avoid the problems associated with it.

3.1 In Favor of Coincidence

Pro-coincidence arguments follow a basic pattern. First, one appeals to a difference in a property at a time during which the objects are co-located

[9] Unfortunately, there is no settled terminology in the literature, although there is an emerging common ground that the sharing of parts at a time is a key factor in the puzzles. (See, e.g. Merricks 2001, Olson 2001, Rea 1997a.) Our use of 'coincidence' is similar to Rea's (1997a) use of 'constitution', with the exception that 'x and y coincide', for us, implies 'x is distinct from y'. Some philosophers (e.g. Olson 2001) use 'material coincidence' for coincidence in the part-sharing sense.

(i.e. located exactly in the same spatial region). Then, by an application of Leibniz's Law, one infers that the relevant objects are distinct. In the statue/lump fission case, one might claim that during the time of co-location the lump has the property of *being such that it will survive the flattening*, but the statue does not. In the fusion case, one might claim that during co-location the lump has the property of *having been shapeless*, but the statue does not. Finally, one argues that the objects coincide while co-located.

One can resist such arguments by denying either the case for distinctness or the case for coincidence given distinctness. We will very briefly discuss the prospects for these paths of resistance. Our goal is only to show the difficulty of blocking the case for coincidence.

Resisting the case for distinctness involves taking one of three options: denying the existence of one (or both) of the involved objects; denying that they differ in a property; or denying the application of Leibniz's Law. Each will be a hard sell. Perhaps it is not especially hard to deny the existence of statues or lumps (students have no trouble). But statue/lump puzzles are not, fundamentally, about just statues and lumps. They are about instantiations of two kind properties, one related more intimately to an object's form, the other to its matter. Thus, for example, there are statue/lump puzzles about persons and bodies. Here eliminativism begins to seem like overkill: ''so, you're telling me that I have to believe that I don't exist or that my body doesn't exist, just to answer these silly puzzles!''[10] Denying difference in a property will also be difficult, for it will require denying that statues and lumps have the persistence conditions associated with ''statue'' and ''lump'' (Burke 1994). How could a statue pre-date its so-called creation, or post-date its so-called destruction? Finally, denying the application of Leibniz's Law is a radical solution, even if not out of the question.

The other way to block pro-coincidence arguments is to deny coincidence given distinctness. One might claim that, while co-located, the statue has an *arm* (say) among its parts, while the lump does not, and so the two do not coincide. However, the statue and the lump are composed of the same atoms, and so any part of one of them will consist of atoms all of which are parts of the other. If certain atoms, which are parts of the

[10] See also McGrath (2005).

lump, compose the statue's arm, then why isn't the arm part of the lump, too?[11] Another hard sell.[12]

3.2 *Against Coincidence*

On the other side, there are powerful reasons to deny coincidence in the statue/lump puzzles above, and in all the puzzles, but articulating them requires some care.[13] Here we put aside arguments that concede distinctness of co-located objects but reject the case for coincidence. The argument that we will discuss presupposes that if distinct objects are co-located, then they must also coincide.

It will help to work backward. Pro-coincidence arguments depend essentially on an assertion that the objects involved differ in a property. But perhaps, although such assertions are intuitive (and follow from plausible principles), they are an opponent's best target (when compared to Leibniz's Law and eliminativism, e.g.). If an opponent could argue convincingly for supervenience principles of the form, *if x and y have the same parts at t, then x and y cannot differ in such and such respects at t*, where the relevant respects include the very respects the pro-coincidence

[11] For further discussion of this move, see Wasserman (2002). Another argument against coincidence, granting distinctness, might be based on the mereological principle called *weak supplementation*: if x is a proper part of y, then there is some z such that z is part of y but is discrete from x (cf. Simons 1987). It will be necessary to relativize this to times, for our purposes: if x is a proper part of y at t, then there is some z that is a part of y at t but is discrete from x at t. Suppose Statue isn't Lump, but that they have all the same parts at t. Then Statue is a part of Lump at t but is distinct from Lump, and so is a proper part of Lump at t. And yet there is no part of Lump at t that is discrete from Statue at t. However, it is unclear why the coincidence theorist should not deny the truth of *weak supplementation* for the temporal mereology of material objects. Given a suitable extensional theory of spatial regions, the following will do much of the work of *weak supplementation* in the temporal mereology of material objects: if x is a proper part of y at t, and x's spatial region at t is a proper subregion of y's at t, then y has at t a part z which is discrete from x at t.

[12] Not only is this a hard sell, but it does nothing to touch anti-coincidence arguments that are based on the weaker assumption that the statue and the lump share all parts at some level of decomposition, e.g. at the molecular or atomic level.

[13] An initial thought is that there is something incoherent in the idea of two things being co-located. But, as some philosophers have pointed out (e.g. Merricks 2001, Olson 2001), co-location between distinct entities of different ontological categories is not particularly objectionable: events and their subjects may be co-located, as may a surface and a shape trope. (One entity would derive its location from the other, in virtue of being dependent on it.) And perhaps even two objects could be co-located, so long as they were made up of fundamentally different but "interpenetrating" atoms or bits of matter.

arguments depend on, then the opponent would take the wind out of the coincidence theorist's sails. More than this, if the class of relevant respects were extensive enough, the case for identity would become irresistible.

We need to introduce some terminology for various kinds of properties.[14] Following Perry (1972: 470), let us call a property *basic* iff its possession at a moment of time t depends on events happening at t but not on any events happening before or after t. Basic properties need not be intrinsic, but they are, as Chisholm (1976) puts it, *intrinsic to a time*. Thus, being located five feet from something is a basic property, though it is extrinsic. We also want to define *futural* and *historical* properties. Here some care is needed. We want the property ''will attend class next week'', for example, to count as futural. However, its possession at a time t will arguably depend on relations holding between events at t and events a week later than t (e.g. relations of spatiotemporal continuity). So, we should not define futural properties as those properties that depend only on events happening after t, but rather as those depending on events after t but not depending on events before t. We define *historical* properties, similarly, as those that depend for their possession at t on events before t but not on events after t.[15]

These categories of property are mutually exclusive but not jointly exhaustive (consider the property *being the first-born son of a woman who will climb Everest*). However, an object's basic, futural, and historical properties jointly fix all of its properties (recall that we are ignoring modal properties).

Return to the two statue/lump cases. One sort of argument for distinctness, applicable to both cases, appeals to basic differences at the time of co-location (e.g. the statue is beautiful; the lump isn't). Another

[14] In this chapter, I reserve the word 'property' for properties that are (i) qualitative, and (ii) not built up even in part from ''zero-place'' properties. An atomic qualitative property, roughly, is one that does not constitutively involve a particular individual. Complex qualitative properties are those that are built up from atomic qualitative properties. (This rough characterization of qualitative properties is borrowed from Adams 1979.) A ''zero-place'' property is one that, intuitively, does not impose on its possessors a real condition for its possession. A mark of zero-place properties is that they are expressible (either in English or in some extension of it) by a predicate of the form 'is such that p', where 'p' is an eternal sentence (see Zalta 1983).

I limit the use of 'property' in this way in order to avoid having regularly to note irrelevant exceptions to general principles about the sharing of properties.

[15] A similar classification is available for relations. We will speak of basic relations below.

sort of argument appeals to futural differences in the fission case and historical differences in the fusion case. The anti-coincidence argument below thus has three steps. *Step One* rules out basic differences among coincidents; *Step Two* rules out futural differences; and *Step Three* historical differences.[16]

Step One. The first step appeals to the principle that in all (nomologically) possible worlds, if objects x and y have all the same parts at a time, then x and y have all the same basic properties at that time. We abbreviate this as *Same Parts ⇒ Same Basics*. This principle may be defended as follows. Perhaps objects could and even do coincide. But there are certain properties that coinciding objects must share while coincident. Sphericality, for example, is fixed by having parts interrelated in certain ways, and so coincidents must be alike in this property. And the same goes for all *material* intrinsic basic properties. But it's also impossible for there to be coincident objects that differ in material *extrinsic* basic properties, such as being five feet away from a third object. And yet won't an object's having the property of beauty, a putatively non-material basic property, be fixed by the totality of its basic material properties? In fact, won't all an object's basic properties be so fixed?[17]

[16] I do not claim originality for all three steps of the argument to follow, particularly not the first. Arguments similar to it in certain ways have been put forward by Burke (1992), Heller (1990), Olson (2001), van Inwagen (1990b), and Zimmerman (1995), among others.

[17] *Same Parts ⇒ Same Basics* will be resisted by some coincidence theorists, including Baker (2000), Johnston (1992), and Wiggins (1980). Rea (1997b) suggests how a coincidence theorist might reply to objections based on this principle. Coinciding objects share all parts at a time, but they supervene on different facts about how those parts (or some subclass of them) are arranged at that time. The statue supervenes on certain things being arranged *statuewise*, and the lump supervenes on certain things being arranged *lumpwise*. This is just what it is to be a statue or a lump. An object will be a candidate for those properties that are appropriate to its supervenience base and not a candidate for those inappropriate to that base. The statue will therefore count as beautiful, but the lump will not; the person will count as thinking, but the body will not. The principle *Same Parts ⇒ Same Basics* will then have to be denied. Beauty is a basic property (let's assume), and so despite having all the same parts, the statue and the lump are not basic duplicates. (Side point: Rea's strategy might be used to argue that the statue and the lump in fact do *not* have all the same parts, despite having all the same atomic parts. Perhaps the statue, in virtue of its supervenience base, is eligible to have an arm as a part, while the lump, in virtue of its base, is not.)

Olson (2001) rightly notes that Rea's proposal leaves unexplained, concerning the object which is in fact beautiful, why *it* is beautiful but the object it coincides with isn't. Both are composed of the same things, and so the same things arranged in exactly the same

However, *Step One* is not enough. The coincidence theorist might concede that the statue and the lump, while co-located, are basic duplicates, but insist that they are not duplicates with respect to their non-basic properties, and therefore must be distinct.[18]

Step Two. This step, and the next, both depend on the crucial idea of a kind of *de re* lawful determination of an object's properties, an idea which will be central to this chapter. We will assume that laws—at least laws in the actual world—subsume objects at a time by virtue of their properties. Rather than determination *simpliciter*, then, we will speak of objects as being determined to be F at a time *given* that it is G at that time. Here, then, is the general scheme for lawful determination:

> *Lawful Determination*:
> O is lawfully determined to be F at t given that it is G at t iff: O is G at t and the generalization *All Gs at a time are Fs at that time* is nomologically necessary.

Hopefully, the core idea is clear enough for now. It will become clearer as we proceed.

Now for *Step Two*. Intrinsic duplicates at a time need not have all the same futural properties. *Basic* duplicates, however, are not only intrinsic duplicates; they are also duplicates with respect to all environment-relating properties. To the extent that an object's futural properties at a time are lawfully determined given its basic *and* historical properties, it seems that its futural properties should be lawfully determined given its basic properties alone. That is, the relevant law-grounded

ways. Why does beauty apply to one but not the other? Why indeed does *being a statue* apply to one rather than the other? Doubts about the answerability of such questions are exactly what motivate *Same Parts ⇒ Same Basics*.

I devote so little space to discussion of *Step One* in the body of the chapter because the standard 4Dist treatment of the puzzles concedes this step, as we shall see. I touch on Rea's proposal again in my extended note 2.

18 An aside: if the coincidence theorist accepts *Same Parts ⇒ Same Basics*, then she will be hard-pressed to allow for permanent coincidence. Permanently coinciding objects would share all their basic properties at every moment of their existence. In addition, because they have the same parts at all times, those parts at those times would stand in all the same ''unity'' relations (spatiotemporal continuity, etc.). It would apparently follow that they share all the same futural and historical properties. What would make the two objects distinct, then? Not qualitative difference (they are never qualitatively different). Not spatial separation (they are never spatially separated). Their distinctness would have to be grounded in modal differences which themselves would have to float free of the non-modal properties and compositional histories of the objects.

generalization should be of the form *All Bs at a time are Fs at that time*, where 'B' picks out a possibly complex basic property. But at every moment of its existence, an object's futural properties are determined given its basic and historical properties. It follows that (in nomologically possible worlds) basic duplicates at a time are futural duplicates at that time, or in other words: *Same Basics ⇒ Same Futurals*. This is the distinctive principle of *Step Two*. Putting it together with *Same Parts ⇒ Same Basics* from *Step One*, we arrive at *Same Parts ⇒Same Futurals*.

Before looking further into this argument, note that *Same Parts ⇒Same Futurals* clearly rules out futural differences in the statue/lump puzzle: given that the statue and lump share all parts at t, and so are basic duplicates at t, they must also be futural duplicates at t, contra the standard pro-coincidence argument. In fact, the principle rules out fission generally, at least in nonsymmetrical nomologically possible worlds.[19]

What about that key move of *Step Two*, that is, the claim that objects' futural properties are lawfully determined at every moment of their existence given their basic properties? This claim is not empirically unrevisable, nor is it intended to be. Properly interpreted, our best scientific theories may turn out to undermine the claim that an object's futural properties are deterministically fixed by the laws, given their basic properties. Perhaps only *probabilities* for having various futural properties are so determined. This would not be too much of a blow to the defender of *Step Two*. She could appeal, instead, to the principle *Same Basics ⇒ Same Probabilities for Futurals*, noting that in the standard statue/lump fission case, the statue has a low and the lump a

[19] The qualification about nonsymmetrical worlds needs explanation. A nonsymmetrical world, let us stipulate, is one in which objects that are non-coincident at a time basically differ at that time. Now even in symmetrical worlds, if objects that coincide at t differ in how long they will survive after a time t, then they will differ in a futural property at t: *being an object that will exist for such and such length of time*. This is why all statue/lump cases are ruled out by *Same Parts ⇒ Same Futurals*. However, suppose x and y are coincident at t but also survive for exactly the same length of time after t. Then if they inhabit a symmetrical world, they could be futural duplicates at t. However, if we limit our consideration to nonsymmetrical nomologically possible worlds, then *Same Parts ⇒ Same Futurals* is sufficient to rule out fission. This is because, assuming x and y coincide at t_1, and that they exist but do not coincide at $t_2 (> t_1)$, then they will differ in some basic property B at t_2, and so will differ in a futural property at t_1: *being an object that will be B at the time that is such and such distance in the future*.

high probability of surviving the flattening blow, even though they are basic duplicates at the time. However, perhaps historical properties in some cases matter even to the determination of probabilities for futural properties. *Step Two* might then be revised appropriately, by appealing, say, to (*Same Historicals & Same Basics*) ⇒ *Same Probabilities of Futurals*. This principle will undermine futural differences in at least those statue/lump cases in which the statue and the lump have been coincident throughout their history. And so the cycle of possible refutations and possible revisions continues.[20]

But what if our best theories show decisively that there is simply no *de re* lawful determination of any kind remotely relevant to the puzzles of coincidence? This cannot be ruled out a priori. Still, for most of us, it would take the earnest testimony of the scientific community for us to give up beliefs like *this* one: the last time you released the apple from midair, it didn't just happen to fall, it *had* to fall given its situation at the time of release. Certainly, most of us are not prepared to give up such beliefs simply because holding onto them makes it hard to solve the puzzles of coincidence. And obviously the releasing of an apple is not a special case. We therefore seem to have good defeasible reason to accept *Same Basics* ⇒ *Same Futurals*.

Step Three. Finally, we need a third step to rule out historical differences between coincidents. One might hope to argue that the laws determine an object's historical properties at a time given its basic properties at that time, and then give an argument exactly analogous to *Step Two* for *Same Basics* ⇒ *Same Historicals*. I will not explore this possibility, except to remark that there is some reason to think that, despite that principle's lack of intuitive appeal, an examination of the proposed laws in our best scientific theories may provide just as much support for *Same Basics* ⇒ *Same Historicals* as for *Same Basics* ⇒ *Same Futurals* (Hoefer 2004).

[20] Recall that we are using 'property' in such a way as to exclude zero-place properties. This exclusion is relevant here. If our best science allows for the possibility of certain sorts of objects going into or out of existence *ex nihilo*, under special conditions (e.g. white holes), then if there is such an event in the future, all of us right now have a zero-place future-oriented ''property'' that isn't lawfully determined given our current basic properties (nor perhaps is its probability for us lawfully determined either). Because we are putting aside zero-place properties and ones built up from them, the anti-coincidence theorist needn't worry about *that* kind of empirical refutation.

However, our *Step Three* proceeds differently. It relies on *Step Two* to rule out fusion, and then having ruled out fusion, rules out historical differences. Fusion is ruled out by appealing to two premises: (1) objects that have undergone fusion either do or at least *could* undergo fission at some later time, but (2) fission is impossible by *Step Two*. If fission is impossible, then co-located objects cannot differ in their historical properties. Thus, *Step Three* gives us *Same Parts* ⇒ *Same Historicals*.[21]

Assuming *Step Two* is sound, (2) is true. Premise (1) can be defended by appeal to combinatorial ideas. If a process produces fusion, it should be possible for it (or a process of its type) to occur later in reverse, producing fission. If the artist's shaping of the lump (plus a suitable creative intention) brings a statue into existence, then the artist's unshaping of the lump (perhaps plus a suitable destructive intention) should destroy it. If cutting off the cat Tibbles' tail produces fusion between Tibbles and Tib (one of Tibbles' previously proper parts), then reconnecting it should produce fission between the two. And so on.[22]

All three steps of the anti-coincidence argument are now in place. Together they rule out both fission and fusion in all the puzzles.

In this section, I have taken pains to make it clear, not only why coincidence is hard to deny, but also and especially why coincidence is so problematic. In the next section, as we examine the standard 4Dist treatment of the puzzles, the question of overriding importance to us will be whether the 4Dist, like the 3Dist, must wrestle with the anti-coincidence argument we have given or whether she can somehow sidestep it.

[21] Strictly speaking, we need to add a qualification about symmetrical worlds like the one discussed in note 19. I let this pass, to keep things simple.

[22] Note that (1) is based on combinatorial ideas that are neutral between counterpart theory and its principal competitors. The counterpart theorist will merely say that in the possible world in which the fission occurs, it occurs not between the actually fused objects but between their counterparts in that world.

One might wonder why, in place of my *Step Three*, I did not simply invoke the intuitive mysteriousness of how coincidents at a time could differ historically (Olson 2001). However, as Perry (1972) pointed out, objects sharing all parts at a region of space can differ in their spatially non-basic properties. He gives the example of Siamese twins sharing a hand. Not that the analogy *proves* anything, but it bolsters the anti-coincidence theorist's case if she can give a reason for thinking that there is an obstacle to recognizing historical differences among coincidents which lacks a parallel in the spatial case. *Step Three* provides such a reason. We will return to these issues in Section 4.

4. CAN 4DISM HELP?

4.1 An Easy and Painless Answer?

We have seen that, putting aside the question of 4Dism, there is no easy and painless answer to the fission or fusion puzzles. Now see things in 4D. Many 4Dists will predict that the problem will look *much* less worrisome. In fact, 4Dists commonly talk as if the puzzles simply disappear once we adopt 4Dism. The idea, touched on earlier, is based on the familiar mantra ''partial overlap is not coincidence''. What is commonly called temporary coincidence, given 4Dism, turns out to be just a kind of partial overlap, in particular partial *temporal* overlap, that is, the sharing of some but not all temporal parts. Only total overlap deserves to be called coincidence, for only with total overlap do things share all parts. Partial overlap is metaphysically unproblematic on its face (obviously, roads can partially overlap and be distinct, and similarly for temporally extended objects). Under 4Dism, there is not even a prima facie threat of genuine coincidence, but only partial overlap, and so the problem is avoided.

The standard 3Dist response is to turn to the modal variants. For example, the 3Dist will refer to Gibbard's (1975) modal variant of the statue/lump puzzle. Suppose a lump (Lumpl) and a statue (Goliath) have the same spatiotemporal career, and so, under 4Dism, have the same 4D extent. Still, they have different modal properties, because one *could* have survived flattening while the other *couldn't* and so are distinct. Even the 4Dist must admit that this is coincidence.[23]

To avoid recognizing coincidence, 4Dists standardly appeal to what Harold Noonan (1991, 1992) calls ''the Abelardian view of modal predicates''. An Abelardian predicate is one ''whose reference can be affected by the subject term to which it is attached'' (1992: 134).[24] A good example comes from Quine: 'Giorgione is so-called because of size.' The predicate refers to or expresses different properties depending on which name it is attached to. Abelardianism makes it possible for the 4Dist to affirm the truth of two seemingly incompatible statements:

[23] See van Inwagen (1990a) for a modal version of the Tibbles or Body-Minus puzzle.

[24] There are different sorts of Abelardian views about modal predicates: one might take it as an irreducible fact that modal predication is relative to a sortal that is contextually supplied by the subject term, or one might try to understand such predication in terms of inconstant counterpart theory (cf. Lewis 1986). See Fine (2003) for a discussion and evaluation of both kinds of proposal.

(A) Lumpl could have survived flattening but Goliath couldn't have.
(B) Lumpl is Goliath.

For, (A) is true iff, relative to the sortal "lump of clay" Lumpl could have survived flattening, and relative to the sortal "statue" Goliath couldn't have survived flattening. (A) is therefore compatible with (B).

The Abelardian move is available to the 3Dist, too. But if, as it is supposed, the 3D coincidence theorist is already burdened with genuine coincidence in the original puzzles, there is little to be gained from Abelardianism in the modal variants. Not so, for the 4Dist. She does fine by the originals and only encounters problems in their modal variants, because, given 4Dism, only in the latter is there genuine coincidence. 4Dism therefore seems to have an edge over 3Dism on the original puzzles.[25]

4.2 Problem Sets and 4D Translation

Before we try to determine whether the 4Dist does have such an easy way with the original puzzles of coincidence, we need to say more about how exactly the 4Dist is bringing 4Dism to bear on these puzzles. This is the goal of the present section.

Let us introduce the idea of a problem set. A problem set is a set of apparently inconsistent statements, each of which appears true. Puzzling cases generate problem sets. For fission puzzles, the most troubling problem set consists of three statements: one asserting sameness of parts at a time t; one asserting a futural difference at t; and one asserting a supervenience principle ruling the possibility of that very sort of difference—in particular, the principle *Same Parts ⇒ Same Futurals*. For fusion puzzles, the most troubling problem set takes an exactly analogous shape, with 'historical' substituted for 'futural'. (I am assuming that problem sets concerning basic differences at t are less troubling.)

A problem set (or the problem represented by the problem set) may be successfully answered in one of three ways: one can dissolve it, by showing how the problem set is consistent after all; one can solve it straightforwardly, by showing that one or more members of the problem set are false; or one can give it a sophisticated solution, by showing

[25] There would be no real edge if Abelardianism about temporal predication had exactly the same advantages and disadvantages as Abelardianism about modal predication, as is claimed by Fine (1980). But for argument's sake, let us assume otherwise.

that one or more members of the problem set is indeterminate in truth-value. (Only the first two sorts of answers will concern us here.) Talk of *avoiding* problems is a bit vague, but we can think of avoiding a problem as having a simple and painless answer to its corresponding problem set. Finally, to the extent that a problem set captures what is puzzling about a case, we can say that answering it suffices for answering the puzzle.

What the 4Dist must do, if she is to answer puzzles like the puzzles of coincidence, is to connect her special 4D talk of temporal parts with the talk of ordinary objects that figures in the statements making up the puzzle's problem set. And she has a natural way of doing this. If ordinary objects are 4D, then, as we briefly noted in Section 2, to avoid unexplained property correlations between ordinary objects and their temporal parts, the 4Dist will want to understand the possession of properties at times by ordinary objects as deriving from facts about those objects' temporal parts. She will want to analyze, or as we will say, to *translate*, statements about ordinary objects having properties at times into statements about those objects having temporal parts with those properties, or suitably related ones.

The 4Dist needs a scheme of translation. For *basic* properties, the scheme is straightforward:

4D Translation Scheme for Basic Properties/Relations
Tr('x is F at t') = 'the t-temporal part of x is F' (or, for short, 'x_t is F')
Tr('$Rx_1, \ldots, x_n$ at t') = 'the t-temporal part of x_1 is R-related to the t-temporal part of $x_2, \ldots$ the t-temporal part of x_n' (or, for short, '$Rx_{1t}, \ldots x_{nt}$').

For non-basic properties, no such simple scheme is available. A non-basic property is a property that applies to an object at a time at least partly in virtue of events happening at other times. The best the 4Dist can say in general is that a 4D object has a non-basic property at a time t in virtue of its having (or lacking) temporal parts with certain basic properties at certain times before, at, and after t. This is hardly worthy of enshrining as a translation scheme for non-basic properties. For particular properties, the 4Dist can be more informative: a person has the property of *awaking* at a time t, for example, in virtue of the fact that her t-temporal part is awake and she has sleeping temporal parts for a suitably long period of time previous to t. So 4D translation of ascriptions of non-basic properties will proceed case by case. However,

and this will be relevant to our concerns, when translating open sentences of the form 'x has P at t', where 'P' ranges over non-basic properties, these sentences may be replaced by 'Tr(P, t) holds of x'. So, for example, the translation of 'for all futural properties F, x has F at t iff y has F at t', will be: 'for all futural properties P, Tr(P, t) holds of x if and only if Tr(P, t) holds of y.'

With these rudimentary elements of the 4D translation scheme in place, we can see that the 4Dist will not be able to *dissolve* the puzzles of coincidence. Consider the statue/lump fission case again. Here is the original problem set:

(Same Parts)
Statue and Lump share all their parts at t.

(Same Parts ⇒ Same Futurals)
For all x, y, if x and y share all their parts at a time t, then for all futural properties F, x has F at t iff y has F at t.

(Futural-Difference)
Statue and Lump differ in a futural property at t.

And here are the translations:

(Tr(*Same Parts*))
Statue_t and Lump_t have all the same parts.

(Tr(*Same Parts ⇒ Same Futurals*))
For all x and y, if x_t and y_t have all the same parts, then for any futural property F, Tr(F, t) holds of x iff Tr(F, t) holds of y.

(Tr(*Futural-Difference*))
There is a futural property F such that Tr(F, t) holds of one but not both of Statue and Lump.

The 4D translations are inconsistent, and so cannot, via the 4D translation scheme, reveal the originals to be consistent. The same goes for all of the puzzles of coincidence.

Nor, I submit, should the 4Dist *want* a scheme of translation that yields a consistent translation set, when applied to problem sets that are inconsistent when interpreted according to their surface logical forms. The 4Dist is not aiming to provide us with a revisionary account of logical form, or the features of ordinary statements that determine

which arguments involving them are valid and which are not. 4D translation is not meant to reveal the true logical form of ordinary statements, but rather *what it is* for the translated statements to be true. By her own lights, then, she should never seek to use 4D translation to dissolve any puzzles whose problem sets are inconsistent when interpreted in accordance with surface logical form. (An Abelardian about temporal predicates, by contrast, is revisionary about logical form.)

The 4Dist, reading along, might protest: ''the puzzles don't threaten coincidence for me, and so can't pose for me the problems uniquely associated with coincidence.'' My response—and this may be all-too-obvious by now—is that the problem posed by the statue/lump case doesn't concern coincidence *simpliciter*, that is, 4D coincidence, but coincidence at a time. The 4Dist *does* accept the notion of coincidence at a time, because she accepts the notion of parthood at a time. (She translates 'x is part of y at t' as 'x_t is part of y_t'.) Furthermore, at least if she appeals to the ''partial overlap'' account, the 4Dist is committed to thinking that the notion of coincidence at a time applies to the relevant objects in the puzzle cases; she will agree that the statement *Same Parts* is true. Consequently, if there are problems inherent in objects coinciding at a time, then the 4Dist inherits those problems.

4.3 A 4D Solution?

So, the 4Dist cannot *dissolve* the puzzles of coincidence. Can she *solve* them? Can she help us see which members of the relevant problem sets are false, and why? We are not interested in solutions that are available with equal plausibility to both 3Dists and 4Dists. We are interested in *distinctively* 4D solutions. Such solutions may be available to 3Dists, but they must be considerably less plausible under 3Dism than under 4Dism. Moreover, we are not interested in distinctively 4D solutions that are easily bested by certain other solutions available to 3Dists. To have an edge over the 3Dist in answering a puzzle, the 4Dist needs a distinctively 4D solution that is more plausible under 4Dism than any of the solutions available to 3Dists are under 3Dism.

A distinctively 4D solution will rely on what distinguishes 4Dism from 3Dism, and therefore on the distinctive 4D ontology and ideology. In *ontology*, the 4Dist accepts a plenum of stages, as well as persisting objects having them as parts, and having fusions of them as parts. The

3Dist doesn't. In *ideology*, as we have seen, the 4Dist accepts the absolute predication of basic properties and relations and analyzes temporally relative predication of such properties/relations in terms of their absolute predication. The 3Dist, at least if she is not a presentist,[26] will deny the truth of all absolute predications of such properties/relations, taking the relativized predications as irreducible. Even the presentist 3Dist must deny the truth of absolute predications of basic properties to non-present entities, and so will deny that the truth of a statement predicating roundness to an object at a non-present time t depends in part on anything's being round *simpliciter*.

Our question, then, is whether, and if so how, the 4Dist can make these differences in ontology and ideology work to her advantage in solving the puzzles of coincidence.

To take an example in which these differences do help, at least arguably, consider the puzzle of change, or of ''temporary intrinsics'' (Lewis 1986). Shapes, we would like to say, are intrinsic properties. But then how can an object change its shape? A change in shape would seem to require it to have two incompatible shape properties. Standard 3Dist solutions appeal to some sort of relationality: perhaps shapes are relation to times, or if not, then at least they are instantiable only relative to a time. The 4Dist agrees that the shapes of persisting objects are had only relative to a time, and in this respect the 4Dist is on all fours with the 3Dist. But she has the resources to accommodate shapes that are not only intrinsic properties but are had *simpliciter*. For, given the 4D translation scheme, the having of shapes by persisting objects at times derives from the having of shapes *simpliciter* by their temporal parts. The 3Dist cannot duplicate this answer, not even the presentist 3Dist. The 4Dist, therefore, has a distinctive solution. Whether it is superior to all 3D solutions is a more difficult question, but perhaps the 4Dist has reason for hope.[27]

[26] The presentist claims that absolute predications of basic properties/relations can be true, but only for present entities presently instantiating those properties/relations. So, the statement ''I am seated'' can be true, without temporal qualification, only if I am presently seated. As mentioned earlier, the presentist will analyze temporally relative predication in terms of absolute predication, as follows: x is F at t iff were t present, x would be F.

[27] Similarly, the 4Dist has a distinctive treatment of the following time-travel puzzle. I travel back to 1975 and take a look at ''my past self''. Am I, in 1975, three feet or six feet tall? I cannot be both. The distinctive 4D treatment is this: I cannot be said to have a height

There are, of course, many puzzles for which the 4D ontology and ideology plainly give the 4Dist no edge whatever. No one would think of invoking 4Dism to answer philosophical puzzles, say, about the nature of color. This is no surprise: the question whether 4Dism is true has a bearing on some philosophical puzzles, but obviously not on all. Our interest is in whether it has a bearing on the puzzles of coincidence, and if so, whether this bearing gives the 4Dist an edge or not.

Return, then, to the puzzles of coincidence. Nothing in 4Dism gives the 4Dist any special reason to deny the relevant claims of futural or historical difference. 4Dism provides no grounds for thinking statues can survive flattenings, that statues but not lumps come into existence with appropriate shaping by an artist, nor that Tib goes out of existence with the cutting of Tibbles' tail, that gradual plank-replacement isn't sufficient for a ship's survival, and so on.[28] Nor does 4Dism give us a special reason to give up the claim that the relevant objects in the puzzles share all parts while co-located. A distinctively 4Dist solution to these puzzles, then, must set its sights on the relevant supervenience principles: *Same Parts⇒Same Futurals* for fission puzzles, and *Same Parts ⇒ Same Historicals* for fusion puzzles. Here, presumably, is where the mantra ''partial overlap is not coincidence'' helps, if it helps at all.

The first thing we need to discuss is how overlap enters the picture at all. The 4D translation of *Same Parts* takes us most of the distance to overlap, but not quite. To plug the hole, the 4Dist needs the assumption that there is no such thing as 4D coincidence, that is, that there is no coincidence *simpliciter*. This will license the inference that, when x and y share all their parts at a time, not only will x_t and y_t share all their parts, but $x_t = y_t$, so that x and y share a t-temporal part. I know of no 4Dists who admit 4D coincidence, and do not expect to learn of any, because 4Dists could hardly gain an edge on 3Dists if they conceded the

in 1975, because my scattered 1975-temporal part has no height, but there is some truth in claiming that I, in 1975, am three feet tall and in claiming that I, in 1975, am six feet tall: my temporal part consists of two person-like stages, one of which is six feet tall and the other three feet tall. The 3Dist cannot duplicate this treatment, because she lacks the ontology.

[28] Any reason that the 4Dist provides to reject one or another of them will be available with equal plausibility to the 3Dist. For example, if the 4Dist wants to utilize Burke's (1994) ''dominant kinds'' approach, according to which possession of a kind property does not assure possession of persistence conditions associated with that kind, the 3Dist may follow suit.

existence of coincidence *simpliciter* of temporal parts in the original temporal versions of the puzzles.[29]

Suppose, then, that the 4Dist embraces ''No 4D coincidence'' and so is positioned to see the puzzle cases as involving temporal overlap. Overlap in a t-temporal part immediately entails, given the 4D translation scheme for basic properties, that the overlapping objects share all the same basic properties at the times of overlap. (And so the 4Dist need not reject *Same Parts ⇒ Same Basics*.) The challenge is to find something in the 4D ontology and ideology that enables us to see how basic duplicates at a time could have different futural or historical properties.

We will discuss three 4D proposals.

4.4 Three 4D Proposals

Proposal #1: appeal to the mantra ''partial overlap is not coincidence''. To repeat, the story is this: temporary coincidence is partial temporal overlap; and partial overlap is metaphysically innocent (who would raise questions about roads partially overlapping?). This is put forward as a simple and painless solution to the puzzles, and so a way of *avoiding* them. If the appeal to partial overlap succeeds, then something must be wrong with the principles *Same Parts ⇒ Same Futurals* and *Same Parts ⇒ Same Historicals* because they rule out something metaphysically innocent. It is not very satisfying, of course, to be told simply that these principles must be wrong. We would like to know how and why those principles go wrong.

Without an explanation of where these principles go wrong, it may seem that Proposal #1 unjustifiably assumes too close an analogy between existence over space and existence over time. In particular,

[29] The 4Dist may be allowed to rely on the denial of 4D coincidence in giving new and improved 4D translations. Thus, the new translation for *Same Parts ⇒ Same Futurals*, for example, will be this: ''For all x and y, if $x_t = y_t$, then, for all futural properties F, Tr(F, t) holds of x iff Tr(F, t) holds of y.''

In absolute mereology, parthood is standardly taken to be a partial ordering, and so a reflexive, antisymmetric, and transitive relation. The first two conditions jointly require that things sharing all their parts are identical. (When we turn to temporal mereology, antisymmetry becomes more controversial. The 4Dist herself will of course deny that if x is a part of y at t and y is a part of x at t, then x must be y, as a check of the 4D translation of this principle will attest.)

while there may well be important analogies between existence over space and existence over time, perhaps including the central 4D analogy, that the two both involve *extension* (and the concomitant having of parts), the recognition of analogies does not preclude the recognition of disanalogies. In fact, the anti-coincidence argument we developed in Section 3.2 points to a significant disanalogy, as I will next explain.[30]

As discussed earlier, myriad humdrum data provide prima facie reason to accept *Same Parts ⇒ Same Futurals*. The apple *had* to fall when released; it was determined to fall, given its total situation at the time of release. If this ''datum'' is wrong, we need to be told how and why. It is on the basis of apparently representative cases like this that *Same Basics ⇒ Same Futurals* is accepted. The latter, together with *Same Parts ⇒ Same Basics*, which the 4Dist accepts, yields *Same Parts ⇒ Same Futurals*.

But now, let us try to find a spatial analogue to *Same Parts ⇒ Same Futurals*. To begin with, we need a notion of spatially basic properties. Let us say that a property is spatially basic if its instantiation at a region R depends only on events happening within R. Next, we need an analogue to futural properties. I am not sure what this would be. But to play along, let it be ''northerly'' properties, that is, properties whose instantiation at R depend on what is going on at R and due north of the northernmost points in R (relative to some privileged reference frame), but not on what is going on south of R. Here is what we arrive at:

A spatial counterpart of *Same Parts ⇒ Same Futurals*:
If x and y have the same parts at a region R, then x and y have the same *Northerly* properties at R.

This is obviously silly. It gains no support whatever from the equally silly spatial counterpart of *Same Basics ⇒ Same Futurals* (i.e. *Same Spatial Basics ⇒ Same Northerlies*). Thus, while there is a prima facie good case for *Same Parts ⇒ Same Futurals*, there is nothing of the sort

[30] The key principle to consider in asking about disanologies is *Same Parts ⇒ Same Futurals*. As we saw earlier, *Same Parts ⇒ Same Historicals* can be defended on its basis, given only the relatively unproblematic assumption that any fusion-creating process could occur in reverse. So we will focus on the former principle.

for its spatial counterpart. Perhaps the appearance of a disanalogy is misleading, but the appearance is plain.[31]

Proposal #2: appeal to the rich 4D ontology. Doesn't the ontology of 4Dism *guarantee* the existence of objects violating *Same Parts ⇒ Same Futurals*? Consider Bush and his current temporal part. They share all the same parts now. But only Bush will survive and only Bush had a past, and he and his current temporal part differ in futural (and historical) properties. Even if we consider only persisting objects, the principle is violated. Consider the part of Bush which consists of his past up to the present, call it Bush*. 4Dists are committed to Bush*, given that they are committed to Bush (given the mereological account of temporal extension). But Bush has a future while Bush* doesn't. More violations follow if the 4Dist accepts the fusion axiom of absolute mereology (i.e. that any set of objects has a fusion *simpliciter*). The principle *Same Parts ⇒ Same Futurals* is simply dead in the water, if 4Dism is true.

I reply: hold on! No 4Dist should say that our ordinary quantifiers are wide open. They are almost always restricted to the objects that are common to 3Dists and 4Dists, to what we have vaguely called *ordinary* objects: statues, lumps, persons, cats, cells, molecules, and so on. So, when the 4Dist translates an ordinary quantified statement, the quantifiers in the resulting translation must be restricted to ordinary objects.[32] What we need from the 4Dist, then, is a story about why the principle is false even when their quantifiers are so restricted.

To make the case for the restricted version of *Same Parts ⇒ Same Futurals*, we need the following notion of *weak* lawful determination of properties:

[31] When this chapter has been presented, I have often encountered the following objection: ''But surely when we see fission and fusion cases in the eyes of the 4Dist, they seem far less troubling: things can partially overlap!'' Eli Hirsch remarked that students' puzzlement about personal identity fission cases diminishes markedly when they are told Lewis's solution. I agree that, intuitively, when we see fission/fusion cases in 4D they seem less problematic, but I think this intuition relies on a default assumption that sharing temporal parts is relevantly analogous to sharing spatial parts. Once we see that the sharing of temporal parts is constrained lawfully in a way that the sharing of spatial parts isn't, the intuition, if it persists, carries less weight.

[32] If we restrict the quantifier in *Same Parts ⇒ Same Futurals* to ordinary objects, then, in order to keep the problem set inconsistent, we must also add a premise to the effect that statues and lumps are ordinary objects. But the addition is unproblematic and so we will neglect it.

> O is weakly lawfully determined to be F at t given that it is G at t iff: O is an ordinary G at t and the generalization *All ordinary Gs at a time are Fs at that time* is nomologically necessary.

If ordinary objects' futural properties are at all times weakly lawfully determined given their basic properties, then it follows that, for all ordinary objects x and y, if x and y are basic duplicates at a time t, then they are futural duplicates at t. The latter is just the restricted version of *Same Basics ⇒ Same Futurals*, which together with the restricted version of *Same Parts ⇒ Same Basics*, entails the restricted version of *Same Parts ⇒ Same Futurals*.

What is hard to see is how the 4D ontology and ideology provide any reason to deny the key move here: that an ordinary object's futural properties at a time are weakly lawfully determined given its basic properties. Of course, the 4Dist wants to solve the puzzles of coincidence, and so may argue that, if the puzzles are to have solutions, weak determination must not occur, or at least must not occur in fission cases. But to argue in *that* way is not to appeal to distinctive features of 4Dism. The 3Dist might well make the same argument, no more and no less convincingly. To repeat something mentioned in Section 3.2: if we are to give up our belief that the apple we released didn't just happen to fall but *had* to fall, given its situation at the time of release, we will need a better reason than that the truth of this belief rules out one of the most attractive solutions of the puzzles of coincidence. We need to be shown what our mistake is.

Proposal #3: appeal to kind-relativity. The 4Dist might try again: ''No, No! I do not deny beliefs like the one about the apple. And I do want to show you the big mistake you're making. Put simply, we normally neglect kind-relativity. Ordinary objects' futural properties are weakly determined given their basic properties, but only relative to a kind.''[33] Proposal #3 thus denies *Same Parts ⇒ Same Futurals* and *Same Parts ⇒ Same Historicals*, but affirms the principles that result by adding *Same Kinds* to their antecedents.

Clearly, if kind properties were themselves basic properties, then appealing to kind-relative determination would not help the 4Dist. The statue and the lump would be of the same kind, and so would be

[33] Kind-relative weak determination is explainable as follows:

> O is weakly lawfully determined in the kind-relative sense to be F at t given that it is G at t iff: there is some kind K such that O is an ordinary K that is G, and the generalization *All ordinary Ks that are G at a time are F at that time* is nomologically necessary.

determined in the kind-relative sense to have the same futural properties. Similarly, if kind properties were historical, the 4Dist would still be in trouble, for then she would have to say that in statue/lump cases in which the statue and the lump share all their past temporal parts, they are of the same kind just before the flattening, and so should have the same futural properties at that time, which they don't. Rather, the 4Dist must insist that the possession of a kind property at any and all times during an object's career depends on events happening before, during, and after that time. And this is a natural view for the 4Dist to take. She does not want to count all of a statue's proper temporal parts (momentary and longer) as statues. Being a statue at a time (or equivalently at all times) is a matter of being a maximal statueishly interrelated fusion of stages (Lewis 1976). Thus, because the statue is a proper temporal part of the lump, the lump cannot be a statue nor can the statue be a lump.

There are at least two problems with the appeal to kind-relativity. Both may have occurred to you. The first is the scope limitation problem: some puzzles of coincidence involve objects of the same kind, and so cannot be regarded by the 4Dist as proper temporal parts of one another (e.g. the ship of Theseus case, and the fission/fusion cases in the literature on personal identity). So, the appeal to kind-relativity will not solve these puzzles.[34] The second, and more serious, problem is that Proposal #3 is not distinctively 4D. In our statement of it none of the distinctive ideology or ontology of 4Dism entered in. The 3Dist, too, may say that weak determination is always kind-relative and that kinds are neither basic nor historical properties.[35]

[34] It is noteworthy that, while 4Dists are typically happy to give an overlap treatment of the ship of Theseus and all the puzzles involving apparently coincident objects of the same kind, one is hard-pressed to find 3Dists who accept the coincidence solution in these cases. My hunch is that 3Dists find it extremely difficult to deny kind-relative determination while 4Dists think the whole problem is avoided by appeal to the mantra ''partial overlap is not coincidence''. For the 3Dist, no story like Rea's (1997b), referring to different supervenience bases for coinciding objects, can be told for coinciding objects of the same kind. For the 4Dist, partial overlap is unproblematic regardless of the kinds of objects involved (two roads can partially overlap, Siamese twins can, etc.).

[35] As Wasserman (2002) notes, it seems no less problematic for the 4Dist to treat kinds as non-basic and non-historical than it is for the 3Dist. Where the 4Dist will say that what temporal parts an object has help to ground its kind, the 3Dist may say that what parts an object has at various times help to ground its kind.

Taking kinds to be grounded in total careers makes it hard to see how kinds can be invoked in explaining why an object at a time has the futural properties it does. If a particular lump is a lump partly because it persists after a flattening, then it is hard to see how its being a lump can be invoked in an explanation of why it persists.

A rather ad hoc fix to Proposal #3 is to relativize weak determination to something guaranteed to solve the problem of scope limitation: *being an ordinary object that does not share a temporal part with another ordinary object at the relevant time*. Even putting aside its ad hoc character, however, this fix solves at most the scope limitation problem. It does not solve the problem of distinctiveness. The 3Dist may, with equal plausibility, relativize weak determination to ordinary objects that are not coincident with other ordinary objects at the relevant time.[36]

Although we cannot rule out the possibility there is some hitherto unexplored 4D proposal that avoids the shortcomings of those we have discussed, we can say that not only do the 4Dists fail to avoid the puzzles of coincidence, but the prospects look dim for a distinctive and plausible 4D solution.[37]

[36] The ad hoc fix amounts to giving up a picture of the world according to which how a thing is qualitatively at a certain time is explainable (deterministically or even probabilistically) in terms of how it was (if it existed) previously. This can be seen by thinking about fission cases involving things of the same kind. Suppose persons P1 and P2 divide at t. P1 ends up in body A and P2 in body B. So, at t, P1, unlike P2, is such that it will be in body A. Bodies A and B will differ qualitatively in numerous ways. Suppose A is 150 lbs, and B 180 lbs. Recall the apple. I release it, and it falls. It *had* to fall given the basic state it was in. How can we explain this? I can say: ''the apple had to fall because it was basically such and such, *and* it was an apple.'' But now how can we explain why P1 had to end up in a 150 lb body? The thing can't be done. I don't intend this to be a devastating consequence. Even if we think that there is only one person before fission, then unless we say that somehow a person can come to have two minds, we are stuck with the choice of either saying (with Chisholm) that the original person inexplicably ended up in the 150 lb body (rather than 180 lb body) or saying that identity is not an essential basis for egoistic concern.

[37] Mark Moyer noted to me that this argument will not succeed unless the supervenience problem I discuss is the main problem with coincidence. He is right. But, he asks, aren't there other problems on which the 4Dist might do better? Consider this argument:

> It just seems bizarre to think that two things could be in the same place at the same time. Unlike the 3Dist, the 4Dist can help us see how it isn't so bizarre: it's part-sharing! The 3Dist must see coincidence as "overcrowding," but not the 4Dist.

I think the 4Dist does not get much mileage from this argument, although I agree that the 4Dist can bring her distinctive 4D ontology to bear here. A couple of comments. First, it is still bizarre, after we accept 4Dism, to see the statue/lump case as involving two persisting objects, and very bizarre to see personal identity fission cases as involving two persisting persons—even if there is a temporal part in common. Secondly, for the 3Dist, ordinary objects exist exactly in certain temporally extended regions of spacetime. If we had two ordinary objects existing exactly in the same such region, that would be overcrowding; but we don't have this, at least in the temporal versions of the puzzles of coincidence. When the 4Dist pins the charge of ''overcrowding'' on the 3Dist, she is refusing to look beyond the moments of coincidence. (It must be conceded that the language of being wholly present encourages the overcrowding charge. See extended note 1 for a discussion of the appropriateness of that language for formulating 3Dism.)

5. AN ARGUMENT AGAINST 4DISM?

I conclude with some admittedly speculative thoughts. Reflection on lawful determination of properties, I want to suggest, provides a *prima facie* reason to reject 4Dism, or if it does not quite succeed in doing *that*, it does at least help to articulate what 3Dists find so very troubling about 4Dism.

As we noted earlier, the 3Dist should not accept a plenum of stages because such a plenum leads to property correlations that demand explanation in terms of mereologically grounded property-borrowing of the 4D kind. Sider's (2001) ''argument from vagueness'' for 4Dism relies on this very fact. His strategy is, very briefly, as follows. First, argue for universalism about diachronic composition on the basis of vagueness considerations. Briefly: diachronic composition cannot be vague, nor can it be metaphysically arbitrary whether it occurs, and yet it does occur, and therefore it must be universal. Secondly, from diachronic universalism argue, in effect, that there is a plenum of stages. And thirdly, argue from the existence of a plenum of stages to 4Dism.[38]

The argument from vagueness is one of the most promising arguments for 4Dism. But acceptance of its main intermediate conclusion, diachronic universalism, comes at a very high cost. It commits one to all Hirsch-style fusions constructible from a base of ordinary objects (Hirsch 1982). So, consider the Hirsch-style object, *US President*, which coincides with Washington when he is president, Adams when he is president, and so on for all past, present, and future US Presidents. It is not surprising that ordinary thought struggles to make sense of *US President* and its kin: assuming they exist, these are objects whose existence over time is lawless—nothing about their futures is fixed by the laws given their basic and historical properties, even in the

[38] Here are more details. Sider argues, on the basis of vagueness considerations, for what he calls *universalism about minimal diachronic fusion*: for any function f from times to sets of objects, there is some object y that exists at and only at the times in f's domain and which is such that, for any <t,X> in f, y is a fusion at t of X. If this doctrine is true, then, for every object O and time t at which O exists, there is a stage that is a fusion of {O} at t. Given Sider's definition of an instantaneous temporal part, it follows that every object has an instantaneous temporal part at every moment of its existence, which is just how he defines 4Dism. See note 8 for a discussion of whether Sider's argument establishes something worthy of the label ''4Dism''.

kind-relative sense.[39] 4Dists therefore seem to be in the following unhappy dialectical position: what is arguably the best argument in favor of their view commits them to unacceptable objects.[40]

Even without diachronic universalism, the 4Dist is stuck with the likes of Bush* (the proper temporal part of Bush that shares a past with Bush but that does not exist after the present moment). Objects like Bush* may not deserve to be classified with the likes of *US President*, but they still have Hirsch-style features. Consider Bush**, which ends a second after Bush* ends. Bush* is a thick temporal part of both Bush and Bush**. Why does Bush** survive but Bush* perish after Bush*'s last moments? The two are basic-cum-historical duplicates at the time, and seem to be of the same kind to boot.[41] There is no explanation of this difference in survival via the laws. Nor will the spatial analogy help: there is no parallel to lawful dynamics for existence over space.[42]

The 4Dist will probably be unimpressed. Isn't it good enough, she might protest, to respect the lawful dynamics of *ordinary* objects, the sort of objects whose dynamics our scientific theories are used to describe? On this task, she will insist, she performs every bit as well as the 3Dist. However, the 4Dist's commitment to the claim that within the temporal boundaries of any ordinary object there reside countless objects whose dynamics are less than fully lawful will reassure 3Dists: "*That* is why I am 3Dist!" To paraphrase J. J. Thomson: objects coming and going lawlessly is a crazy metaphysic![43]

[39] It is hard to know what to say about the kind of *US President*, but I cannot see why its kind should be different from that of any other Hirsch-style object.

[40] After writing this chapter, I learned of Balashov's (2005), in which it is argued that the 3Dist may block Sider's vagueness argument by appealing to considerations of lawful dynamics. Roughly, the idea is that the 3Dist can draw a non-arbitrary line at minimal D-fusions which violate the laws of motion.

[41] Could the 4Dist claim that "thick" proper temporal parts of ordinary objects such as Bush* and Bush** do not inherit basic properties from stages in the way that ordinary objects do? This question is raised and examined by Katherine Hawley (2001). It is hard to see on what basis the 4Dist could deny such inheritance without denying that these thick proper temporal parts have stages as parts. But if thick proper temporal parts don't have stages as parts then in virtue of what do they persist? They would not perdure or even minimally perdure in the sense of note 6.

[42] See the extended note 2 for an important objection to this argument.

[43] Special thanks go to Jeremy Fantl, Mark Heller, Mark Moyer, Michael Rea, Peter Vallentyne, and an anonymous referee, all of whom provided me extensive comments on versions of this chapter. I also thank Michael Della Rocca, Shelly Kagan, Jon Kvanvig, Andrew Melynk, Ted Poston, Ted Sider, Ryan Wasserman, and Paul Weirich.

EXTENDED NOTES

Extended Note 1. Not all endurantists who accept the mereological account of what it is for an object to have temporal extent will accept the ''wholly present'' formulation of endurantism. In particular, eternalist endurantists will not. Let me explain.

An object is wholly present (WP) at each moment of its existence iff *all its parts exist at each such moment*. Either 'parts' here has the absolutist reading or there is some hidden temporal relativization. Suppose an endurantist accepts the mereological account of temporal extension. Then she will not believe that objects are temporally extended (because if they are they perdure). So, for her, the only way to read 'parts' absolutely would be the presentist way. (Explanation: the presentist holds that temporally relativized predications of the form 'x is F at t', including mereological ones, are reducible and so equivalent to conditionals in which 'F' has no temporal parameter: 'were t present, x would be F' (Merricks 1994). From the perspective of presentism, then, one can reason thus: an object O exists at t just in case, were t present, O would exist; and so just in case, were t present, all of O's parts would exist, and so just in case all of O's parts exist at t.)

However, our endurantist needn't be a presentist. If she is an eternalist, then she will deny that objects have parts *simpliciter*. She will therefore regard it as vacuously true to say that all of a persisting object's parts *simpliciter* exist when it does. But persistence doesn't consist in the satisfaction of a vacuous condition, because not every object persists over every period of time! So reading 'parts' absolutely in WP, the WP formulation of endurantism will be unacceptable to eternalist endurantists who accept the mereological account of temporal extension.

Suppose, on the other hand, that 'parts' has a hidden temporal relativization. Then our endurantist will reject the WP formulation on the grounds that the WP condition is either vacuous or satisfiable only by objects that never change their parts. The reasoning here is as follows (borrowed from Sider 2001: 63–8). There are two main candidates for the interpretation of 'parts' in the WP condition—''parts the object has at *every* time of its existence'' and ''parts the object has at *some* time or other of its existence''. The WP condition is vacuous on the first interpretation, and on the second, it is satisfied only by objects that never change in their parts.

I should finally mention another ''more positive'' account of endurance, which will serve all 3Dists equally well: objects persist and do so by having parts at each moment of their existence, where this having of parts at a time is not reducible to the having of temporal parts.

Extended note 2. Here is an important objection to our argument against 4Dism in Section 5. ''Look, even if existence over space is different in the way you suggest from existence over time, the principles for diachronic composition that lead the 4Dist to accept arbitrary undetached temporal parts and arbitrary diachronic fusions have synchronic analogues that lead the 3Dist to accept arbitrary undetached spatial parts and arbitrary synchronic fusions. But now: is the temporally scattered object *US President* any more mysterious than the spatially scattered object which is the fusion of Bush and the Sears Tower? Is Bush* any more mysterious than the current proper spatial part of Bush that includes everything except his left hand?''

First off, I agree that all these objects would be mysterious. But the issue I am raising in the body of the chapter is about lawful dynamics. Temporally scattered objects and arbitrary undetached temporal parts lack lawful dynamics. However, there is no *simple* argument for thinking that the same goes for their spatial counterparts. In the temporal case, we are adding things to our ontology that share temporal parts, and so which are basic duplicates at certain times without being futural duplicates at those times. By contrast, in the spatial case, we are merely adding more things to our ontology that do not coincide with other things and so do not share their basic properties with anything else. Objects that do not coincide at a time — even strange ones— will differ in basic properties of the form *having parts with such and such basic properties and standing in such and such basic interrelations*. The fusion of Bush and the Sears will therefore not have the same basic properties as anything else (assuming it is not coincident with anything else and assuming our world is nonsymmetrical). So long as coincidence is avoided, *Same Basics* ⇒ *Same Futurals* may be preserved.

The hard part is to say something informative about the basic and kind properties that these spatially odd objects have *without starting down the pathway that leads to accepting coincidence*. Suppose we say that the fusion of Bush and the Sears has the persistence conditions of a ''mere sum'', i.e. it persists iff both Bush and the Sears persist. Why, then, isn't there also a mere sum of all the atoms currently composing *me*? If there is, trouble seems to follow: I persist through the destruction of one of my atoms but the mere sum does not. But wouldn't I be coincident with this mere sum, and so a basic duplicate of it? If so, *Same Basics* ⇒ *Same Futurals* is not preserved.

One way out for the 3Dist is to deny the mereological principles of arbitrary synchronic fusions and arbitrary undetached spatial parts. Granted, denying the latter means rejecting the mereological account of spatial extension at a time. 3Dism does not entail the truth of these principles, nor do 3D solutions to philosophical puzzles require their truth, nor do the best arguments for 3Dism rely on them. The contrast with 4Dism should be clear. It must be said, however, that denying these mereological principles will require the 3Dist to

answer arguments for them (e.g. Lewis's argument for arbitrary synchronic fusions).

However, the 3Dist might be able to accept arbitrary synchronic fusions and undetached parts without sacrificing *Same Basics* ⇒ *Same Futurals*. Suppose she thinks of the relevant objects as mere sums. Quite possibly, she will have to concede that I and the mere sum of my atoms are distinct objects composed by a single set of atoms. However, she might be able to hold on to *Same Basics* ⇒ *Same Futurals*. She may argue (as we noted in Section 3.2) that things composed of the same atoms (or more generally, sharing a level of decomposition) need not be basic duplicates. I can think, after all, whereas the mere sum of atoms can't. Assuming things composed of the same atoms share all parts, this would mean denying *Same Parts* ⇒ *Same Basics*. Here Rea's (1997b) suggestion may be helpful as a start: despite having all the same parts, a difference in supervenience bases might be claimed to ground a difference in basic properties. Of course, this raises questions about the ground for differences in supervenience bases (exactly analogous questions arise if differences in supervenience bases are grounded in difference in kind). Here the 3Dist has a number of options: (1) nothing; (2) differences in past careers; (3) differences in future careers; (4) differences in the set of possible careers. The options are not particularly appealing, but *Same Basics* ⇒ *Same Futurals* is preserved.

REFERENCES

Adams, R. (1979) ''Primitive Thisness and Primitive Identity'', *Journal of Philosophy* 76: 5–26.

Baker, Lynne Rudder (2000) *Persons and Bodies: A Constitution View* (Cambridge: Cambridge University Press).

Balashov, Yuri (2005) ''On Vagueness, 4D and Diachronic Universalism'', *The Australasian Journal of Philosophy* 83: 523–31.

Burke, Michael (1992) ''Copper Statues and Pieces of Copper: A Challenge to the Standard Account'', *Analysis* 52: 12–17.

——(1994) ''Preserving the Principle of One Object to a Place: A Novel Account of the Relations Among Objects, Sorts, Sortals, and Persistence Conditions'', *Philosophy and Phenomenological Research* 54: 591–624. Reprinted in Rea *Material Coincidence: A Reader*, ed. Michael Rea (1997) (Lanham, MD: Rowman & Littlefield).

Cartwright, Helen Morris (1970) ''Quantities'', *Philosophical Review* 79: 25–42.

Chisholm, Roderick M. (1976) *Person and Object: A Metaphysical Study*. (La Salle, Ill: Open Court Publishing Co.).

Fine, Kit (1980) ''Acts, Events, and Things'', in W. Leinfellner, E. Kraemer, and J. Schank, (eds.) *Language and Ontology: Proceedings of the Sixth International Wittgenstein Symposium*, 97–105.

——(2003) ''The Non-Identity of a Material Thing and its Matter'', *Mind* 112: 195–234.

Gibbard, Allan (1975) ''Contingent Identity'', *Journal of Philosophical Logic* 4: 187–221.

Hawley, Katherine (2001) *How Things Persist* (Oxford: Oxford University Press).

Heller, Mark (1990) *The Ontology of Physical Objects* (New York: Cambridge University Press).

——(2000) ''Temporal Overlap is not Coincidence'', *Monist* 83/3: 362–80.

Hirsch, Eli (1982) *The Concept of Identity* (New York: Oxford University Press).

Hoefer, Carl (2004) ''Causal Determinism'', *The Stanford Encyclopedia of Philosophy (Spring 2004 Edition)*, Edward N. Zalta (ed.).

Johnston, Mark (1992) ''Constitution is not Identity'', *Mind* 101/401: 89–105.

Koslicki, Kathrin (2003) ''The Crooked Path from Vagueness to Four-Dimensionalism'', *Philosophical Studies* 114: 107–34.

Lewis, David (1976) ''Survival and Identity'', in A. O. Rorty (ed.) *The Identities of Persons* (Berkeley, CA: University of California Press), 317–41.

——(1983) *Philosophical Papers*, vol. 1 (New York: Oxford University Press).

——(1986) *On the Plurality of Worlds* (Oxford: Blackwell).

McGrath, Matthew (2005) ''No Objects, No Problem?'', *Australasian Journal of Philosophy* 83: 457-86.

Markosian, Ned (1998) ''Simples'', *Australasian Journal of Philosophy* 76: 213–26.

Merricks, Trenton (1994) ''Endurance and Indiscernibility'', *Journal of Philosophy* 91/4: 165–84.

——(2001) *Persons and Objects* (New York: Oxford University Press).

Noonan, Harold (1989) *Personal Identity* (Routledge).

——(1991) ''Indeterminate, Identity, Contingent Identity and Abelardian Predicates'', *Philosophical Quarterly* 41/163: 183–93.

——(1992) ''Coincidence is Identity'', *Mind* 102/405: 133–46.

Oderberg, Daniel (1996) ''Coincidence Under a Sortal'', *Philosophical Review* 105: 145–71.

Olson, Eric (2001) ''Material Coincidence and the Indiscernibility Problem'', *Philosophical Quarterly* 51: 337–55.

——(2003) ''Lowe's Defence of Constitutionalism'', *Philosophical Quarterly* 53: 92–5.

Parsons, Josh (2000) ''Must a Four-Dimensionalist Believe in Temporal Parts?'' *Monist* 83/3: 399–418.

Perry, John (1972) ''Can the Self Divide?'' *Journal of Philosophy* 69/16: 463–88.
Rea, Michael, (ed.) (1997a) ''Introduction'', in *Material Coincidence: A Reader*, ed. Michael Rea (1997) (Lanham, MD: Rowman & Littlefield), pp. xv–lvii.
——(1997b) ''Supervenience and Co-location'', *American Philosophical Quarterly* 34: 367–75.
——(1998) ''Temporal Parts Unmotivated'', *Philosophical Review* 107/2: 225–60.
Robinson, Denis (1982) ''Reidentifying Matter'', *Philosophical Review* 91/3: 317–41.
Sider, Ted (1996) ''All the World's a Stage'', *Australasian Journal of Philosophy* 74/3: 433–51.
——(1997) ''Four-Dimensionalism'', *Philosophical Review* 106/2: 197–232.
——(2001) *Four-Dimensionalism: An Ontology of Persistence and Time* (New York: Oxford University Press).
——(2003) ''Against Vague Existence'', *Philosophical Studies* 114: 135–46.
Simons, Peter (1987) *Parts: A Study in Ontology* (Oxford: Oxford University Press).
Thomson, J. J. (1983) ''Parthood and Identity across Time'', *Journal of Philosophy* 80: 201–220.
van Inwagen, Peter (1981) ''The Doctrine of Arbitrary Undetached Parts'', *Pacific Philosophical Quarterly* 62: 123–37.
——(1987) ''When are Objects Parts?'' *Philosophical Perspectives* 1 (Atascadero, CA: Ridgeview), 21–47.
——(1990a) ''Four-Dimensional Objects'', *Nous* 24: 245–56.
——(1990b) *Material Beings* (Ithaca: Cornell).
Wasserman, Ryan (2002) ''The Standard Objection to the Standard Account'', *Philosophical Studies* 111: 197–216.
Wiggins, David (1980) *Sameness and Substance* (Cambridge, MA: Harvard).
Zalta, Edward (1983) *Abstract Objects: An Introduction to Axiomatic Metaphysics* (Dordrecht: D. Reidel).
Zimmerman, Dean W. (1995) ''Theories of Masses and Problems of Constitution'', *Philosophical Review* 104: 53–110.

6. Time Travel, Coinciding Objects, and Persistence

Cody Gilmore

1. INTRODUCTION

Endurantism, roughly stated, is the view that material objects lack temporal extent and persist through time by 'enduring'—that is, by being *wholly present* at each moment of their careers. Perdurantism is the opposing view that material objects persist by 'perduring'—that is, by being temporally extended and having different *temporal parts* located at different times.[1] In this paper I offer an argument against perdurantism, one based largely on premises that have been used in arguments against endurantism. Perdurantists can resist the argument, but not, I think, without weakening at least one of the relevant anti-endurantist arguments.[2] In one way or another, then, this chapter is meant to alter the overall debate between endurantists and perdurantists to the benefit of the former.

The heart of the chapter is the presentation of a new type of coincidence puzzle. A coincidence puzzle is an apparent counter-example to the following, widely accepted *anti-coincidence principle*:

[1] This usage of the terms 'endure' and 'perdure' is due to Mark Johnston (1983). Both endurantism and perdurantism should be distinguished from the *stage* view (e.g. Sider 2001, Hawley 2001), which differs from standard perdurantism in identifying ordinary material objects, not with temporally extended sums of temporal parts (worms) as perdurantists do, but with so-called 'stages'—i.e. with the entities that perdurantists call 'instantaneous temporal parts of ordinary material objects'. Stage theorists need not and typically do not deny the existence of worms; rather, they merely deny that tables and human beings and other ordinary things are worms. The argument presented in this chapter is not intended to count against the stage view; and I shall ignore it in what follows. (For discussion and criticism of the view, see Haslanger 2003, who dubs it 'exdurantism'.)

[2] One such anti-endurantist argument is Theodore Sider's argument from time travel (2001: 101–9). For further discussion of Sider's argument, see Davidson (2004), Markosian (2004), and Sider (2004).

> It is impossible for numerically distinct material objects to *coincide*—that is, to be (i) wholly present in exactly the same location[3] and (ii) composed, at some level of decomposition, of all the same parts or all the same matter at the given location.[4]

To *solve* such a puzzle, as I shall use the term, is to show that the case in question does not in fact constitute a genuine counter-example to the principle.[5]

Existing coincidence puzzles can be divided into two types, corresponding to the manner in which they bear upon the endurantism v. perdurantism debate. Puzzles of the first type (involving temporary spatial co-location) can be solved simply by abandoning endurantism in favor of perdurantism, whereas those of the second type (involving career-long spatial co-location) remain equally puzzling on both views. In this paper I show that if backward time travel is possible, then a *third* type of coincidence puzzle arises. Puzzles of this third type confront perdurantists and can be solved simply by shifting to endurantism.

The plan for the chapter is as follows. In Section 2 I introduce some new terminology and show how it applies to the older puzzles. In Section 3 I give two examples of the new type of puzzle. Finally, in Section 4, I present the argument against perdurantism and discuss a number of possible responses.

[3] That is, same spacetime region or same place at one time.

[4] This is a rough statement of the principle. To state the principle in a way that would make it both maximally precise and acceptable to all self-described anti-coincidentalists (who disagree amongst themselves as to (i) the adicity of the fundamental parthood relation and (ii) the definability of the three-place relation in terms of the two-place notion, and vice versa), we would need to make the principle highly disjunctive and conditional. In my opinion, it is not especially difficult to formulate such a principle, but it is tedious and the results are difficult to process at a glance. Since the principle would require so much unpacking and the benefits of seeing it written out are minimal, I will not formulate it here.

Moreover, for the sake of brevity, I shall henceforth speak of coincidence as if it involved nothing more than sharing the same location. Nothing will turn on this: for each of the examples that I shall discuss, it will be obvious that if the relevant objects *share the same location*, then they also *share all the same parts or all the same matter at that location*. (None is merely a case of co-location without matter-sharing—e.g. a case in which two 'ghost' particles freely pass through one another without ever sharing parts or matter. Even died-in-the-wool anti-coincidentalists can happily concede the possibility of ghost particles.)

[5] Another sort of solution, of course, is simply to reject the anti-coincidence principle itself (e.g. Johnston 1992, Fine 2003). I do not wish to suggest that I endorse the principle. For ease of exposition, however, I shall speak in the manner indicated above.

2. NEW TERMINOLOGY AND ITS APPLICATION TO EXISTING PUZZLES

I shall begin by introducing three technical terms. (1) Let us say that spacetime region[6] R is a *location* of object O just in case R exactly contains the whole of O, or, synonymously, just in case O *exactly occupies*[7] or is *wholly present* at R. (2) Let us say that spacetime region R is an *S-region* of object O just in case R corresponds to what we would ordinarily think of as a spatial location of O at some instant in O's career.[8] Only instantaneous spacetime regions can be S-regions of objects. (3) Let us say that spacetime region R is the *path* of object O just in case R is the union of the regions that O exactly occupies. (This entails that if O occupies just one region, then *that* region is O's path.) Intuitively, an object's path is the spacetime region that exactly corresponds to O's complete career or life-history. This will always be the union of the object's S-regions.

Given a complete description of the career of some material object O, philosophers on both sides of the endurance v. perdurance dispute can all agree on two things: they can all agree about which region is O's *path*, and they can all agree about which regions are O's *S-regions*.

[6] Throughout this chapter I presuppose that some form of spacetime substantivalism is true.

[7] Let me say a brief word about my understanding of the notion of being *wholly present* at, or *exactly occupying*, a region. (I use these phrases interchangeably, although I am aware than not everyone will approve of this policy.) I take no stand on the question of whether this notion is primitive or defined, but do not attempt to define it here. I assume that it can be partially characterized by the following principle: (1) necessarily, if entity O exactly occupies spacetime region R, then O has, or has-at-R, the same shape and size as R, and O stands, or stands-at-R, in the same spatiotemporal relations to things as does R. It also seems plausible to suppose that exact occupation obeys a second principle: (2) necessarily, if O is a material object and O exactly occupies each of two distinct spacetime regions R and R*, then some sort of immanent causal relation holds between the contents of R and the contents of R*. I assume that there is nothing obviously false or unintelligible about the claim that a single thing exactly occupies each of several non-intersecting, extended regions but not their union or any of their proper subregions. Once the notion of exact occupation is understood, it can be used to define other, arguably less fundamental, locative notions. For example, 'Spacetime point p *lies within* object O' can be defined as 'p belongs to some region that O exactly occupies', and Josh Parsons's 'x is weakly located at R' (Chapter 7) can be defined as 'x exactly occupies some region that intersects R'. For a fuller characterization of exact occupation and other definitions, see my 2006.

[8] Some may find the following, alternative characterization helpful: R is an S-region of O just in case R is a region that O would exactly occupy if O were an enduring object.

What these philosophers *disagree* about is which regions are O's *locations*. Perdurantists will say that O has only one location[9]—namely, its path—and that none of O's S-regions are locations of O. Endurantists will say that O has many locations—namely, its S-regions—and that O's path is not a location of O.[10]

These terms provide us with a convenient method of classifying coincidence puzzles, both old and new. We can divide the old puzzles into two types. We shall say that O and O* are involved in a *type A situation* just in case: (i) O and O* are numerically distinct material objects, (ii) they share one or more of their S-regions, and (iii) they have different paths. And we shall say that O and O* are involved in a *type B situation* just in case: (i) O and O* are numerically distinct material objects, (ii) they have all of their S-regions in common, and (iii) they share the same path. The objects involved in a type A situation are spatially co-located for at least one instant but do not have perfectly co-extensive life-histories. The objects involved in a type B situation are spatially co-located with each other throughout their careers, from which it follows that they do have perfectly coextensive life-histories.

The possibility of type A situations would force the endurantist, but not the perdurantist, to accept the possibility of distinct, coinciding material objects. To see this, suppose that Dion and Theon[11] are involved in such a situation, that is, that they are distinct material objects that share some of their S-regions but don't have the same total path.

[9] So far as I am aware, Hudson (2001) is the only self-described perdurantist who denies this. On his view, ordinary things such as human beings are multi-located spacetime worms, wholly present at each of many, massively overlapping, four-dimensional spacetime regions.

[10] I suspect that some self-described endurantists, especially Josh Parsons (2000 and ch. 7), would balk at this. (According to Parsons, fundamental particles such as quarks are spatially and temporally extended mereological simples. Parsons counts them as endurers merely because they lack proper temporal parts. Moreover, he defines a notion of being wholly located at a region according to which it is analytic truth that an extended simple is wholly located at each of many regions—specifically, at each of the many subregions of that region containing all and only those points that lie within the simple. So I suspect that Parsons would say that a quark is wholly located at each point in its path, not just at its S-regions.) It seems to me, however, that endurantists *ought* to accept the words that I've put in their mouths and that the vast majority of them *would* accept this if they were to consider the matter in these terms.

[11] Dion is a man who loses his left foot at t. Theon is the large part of Dion that, prior to t, consists of all of Dion but his left foot. Assuming that they both exist before and after the amputation, it seems that Dion and Theon share their *post*-t S-regions but have different (though overlapping) *pre*-t S-regions.

Then, like any pair of objects involved in a type A situation, Dion and Theon will coincide if they endure but not if they perdure. If they *endure*, then each of them is wholly present in each of its S-regions. Since they have at least one S-region in common, there is at least one spacetime region at which they are both wholly present, that is, at which they coincide. If, on the other hand, Dion and Theon *perdure*, then each of them is wholly present only at its own path. And since they do not share a path, there is no spacetime region at which they coincide. This shows that type A situations constitute coincidence puzzles that can be solved simply by abandoning endurantism in favor of perdurantism.

The possibility of type B situations, on the other hand, would force both the endurantist and the perdurantist to accept the possibility of distinct, coinciding material objects.[12] This can be seen by supposing that Lumpl and Goliath[13] are involved in such a situation, that is, that they are distinct material objects that share all of their S-regions and their path. Like any such pair, they will coincide *not only* if they endure, but also if they perdure. If they endure, they coincide at each of their many shared S-regions; if they perdure, they coincide at their one shared path. And of course the point is completely general: type B situations, were they to occur, would constitute coincidence puzzles that would remain *equally* puzzling regardless of which of the two rival views about persistence is correct.

3. THE NEW PUZZLES

So much for existing puzzles; I shall now describe a new type of puzzle. Let us say that O and O* are involved in a *type C situation* just in case: (i) O and O* are numerically distinct material objects, (ii) they have *none* of

[12] According to my definition of 'type B situation', O and O* must be numerically distinct in order to be involved in such a situation. So far as I am aware, all actual perdurantists accept the anti-coincidence principle and so deny the possibility of type B situations. Perdurantists differ amongst themselves as to the best technique for explaining away the apparent differences between the objects that seem to be involved in such situations. See, e.g. Lewis (1971), Gibbard (1975), Heller (1990), and Noonan (1993).

[13] Following Gibbard (1975), we can suppose that Lumpl is a statue-shaped lump of clay, that Goliath is a clay statue, that Lumpl and Goliath are created at the same time and destroyed at the same time, and that they share their constituent matter throughout their lives. Of course, Gibbard himself would deny that Lumpl and Goliath are distinct and so would deny that they're involved in a *type B situation*, given my definition of this term.

their S-regions in common, and (iii) they share the same path. The objects involved in a type C situation are never spatially co-located despite the fact that they trace out the same overall path through space-time. I shall present two apparent examples of this type of situation.

The first case involves a form of backward time travel that is familiar from science fiction stories but receives little serious attention from physicists or philosophers of science. The idea is that something persists through time in the normal way for a while then suddenly disappears, reappearing out of nowhere at an earlier time.[14] We are supposed to think of the time-traveler as jumping discontinuously from the later time to the earlier time, rather than gradually and continuously working his way back.

Suppose that some cell is originally created at the beginning of the year 2000 and that it jumps back in time over and over again, never venturing further back in time than the moment of its original creation, and never progressing beyond the end of the year 2002. The cell's entire career is confined to this three-year interval. Suppose also that the cell never leaves the immediate vicinity of my bathtub. If this cell's trips were structured properly, if it made enough of them, and if it underwent the right sorts of intrinsic changes along the way,[15] the cell might compose some macroscopic object that sits in my bathtub for three years. Indeed, the cell might compose an object that by all appearances is a conscious, intelligent human being, one who exhibits the strange behavior of living in my bathtub, and whose constituent cells seem to pop into and out of existence,[16] but who is otherwise quite normal.

[14] This description of the situation presupposes that we can make out some distinction between external time and the time-traveler's proper (or personal) time, so that whereas the reappearance precedes the disappearance in external time, the order of these events is reversed with respect to the time traveler's proper time. Intuitively, external time corresponds to the prevailing global time order, whereas a thing's proper time is what would be measured by a clock carried along with that thing. For more on this distinction, see Lewis (1976).

[15] Plausibly, Cell counts as a backward time traveler only if his sudden disappearances *cause* his earlier reappearances. To ensure that even a Humean about causation will have grounds for holding that for each sudden *disappearance* D of Cell, D causes some specific *reappearance* R of Cell at an earlier time, we may need to supplement the case as follows: let Cell contain a tiny clock that measures the elapse of Cell's proper (or personal) time and that never returns to the same state. Then a given disappearance D will count as the cause of an earlier reappearance R provided that the clock's state at R indicates that R occurs immediately after D (with respect to Cell's proper time).

[16] As I envision the situation, there are many appearances and disappearances of Cell occurring throughout Tubman's life, at various places within his body.

Let us use the name 'Cell' to refer to the time-traveling cell involved in this case, and let us use the name 'Tubman' to refer to the macroscopic object that is completely composed by Cell throughout the three-year interval. Cell and Tubman trace out the same overall spacetime path over the course of their careers. But, intuitively speaking, they do so in different ways, for their S-regions are quite different. At each moment in the interval in question, Tubman has exactly one spatial location, this location being human-sized and human-shaped. Cell, however, has a great many different spatial locations at each moment in this interval, each of these locations being microscopic and cell-shaped.

The fact that Cell and Tubman have different S-regions entails that they are numerically distinct. However, their distinctness can be argued for in other ways as well. (1) Tubman is conscious, but Cell is not.[17] (2) Unlike Cell, Tubman will never travel backward in time. No one will ever see older and younger versions of him in the same room at once. Thus the case of Cell and Tubman seems to be an example of a type C situation: it seems to involve distinct material objects that share their path but have none of their S-regions in common.[18]

Like any pair of objects involved in such a situation, Cell and Tubman coincide if they perdure but not if they endure. If they perdure, then each of them is singly located: each of them is wholly present only at its

[17] Indeed, we might imagine a variant of this case in which: (i) Cell is replaced by Happy, a tiny sentient creature who is always elated, and (ii) Tubman is replaced by Sad, a macroscopic creature who is depressed throughout his life. Though Happy and Sad share a path, the apparent incompatibility of their mental histories strongly suggests that they're distinct.

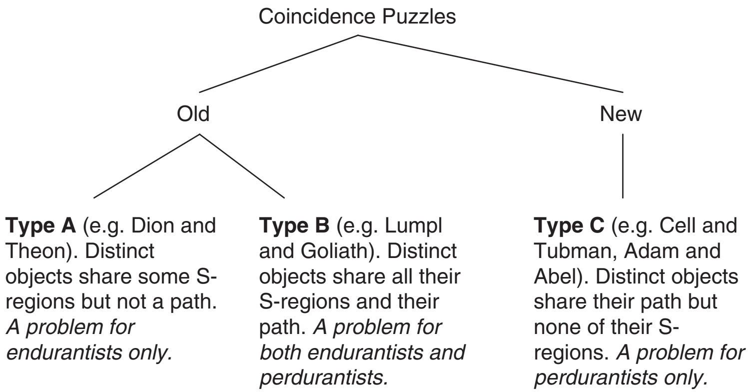
[18] It is worthwhile to contrast the way in which these objects differ from the way in which the objects involved in familiar type B cases (or 'alleged' type B cases) differ. Whereas the latter objects typically differ only with respect to their modal properties, persistence conditions, or sorts, the former objects differ with regard to straightforward physical properties (mass, spatial location, spatial size) and mental properties. Fine (2003) has noted that objects apparently involved in type B situations sometimes seem to differ with respect to not-obviously-modal properties such as *being valuable* or *being damaged*. (A bronze statue, e.g. may be left undamaged by irradiation that damages its constituent piece of bronze.) It seems to me, however, that these differences are in some way *social* or *institutional* and therefore depend on the existence of sentient beings (which, of course, does not by itself entail that these differences are unreal), whereas some of the differences between the objects in my type C cases do not so depend. Thus even if, for some as-yet-unspecified reason, one is inclined to reject arguments for non-identity based on supposed differences with respect to modal properties, persistence conditions, sorts, or properties such as *being valuable* and *being damaged*, one can *still* accept my argument for the non-identity of Cell and Tubman.

entire path. And since they both have the *same* path, they coincide there. If they endure, however, then each of them is multi-located: each of them is wholly present at each of its S-regions (and at those regions only[19]). But since they have *none* of their S-regions in common, there is no spacetime region at which they are both wholly present; that is, they do not coincide anywhere. Rather, Cell is wholly present only at many microscopic, cell-shaped, instantaneous spacetime regions, while Tubman is wholly present only at many macroscopic, human-shaped, instantaneous spacetime regions.

Thus type C situations constitute coincidence puzzles that can be solved simply by shifting from perdurantism to endurantism.[20] The possibility of such situations would force the perdurantist, but not the endurantist, to accept the possibility of distinct, coinciding material objects.

[19] Why shouldn't the endurantist say that Cell is also wholly present at many macroscopic, human-shaped regions, in addition to being wholly present at many microscopic, cell-shaped regions? For at least three reasons: (1) If the endurantist were to say this, he would be admitting that both Cell and Tubman are wholly present at each of the relevant human-shaped regions, hence that they coincide at those regions. Thus endurantism would no longer solve the puzzle in question. (2) As I mentioned in a previous note, being wholly present in a region entails having (at that region) the same shape and size as that region. So if the endurantist were to say that Cell is wholly present at many macroscopic, human-shaped regions, he would be committed to the implausible view that Cell, a mere human cell, is human-sized and human-shaped (at certain regions). (3) Assume endurantism is true. We know, then, that Cell's career begins in some small, cell-shaped region; call it R. (Perhaps Cell has no *first* S-region; still, we know there's some brief series of such regions where Cell's career begins.) So we know that Cell is wholly present at R (or at each member of the series). We also know, it seems to me, that if a material object is wholly present at each of two different regions, R and R*, then there must be the appropriate sort of immanent causal relation between R and R* (or their 'contents'). But the small, cell-shaped region R, it seems, does not bear the right sort of immanent causal relation to any large, human-shaped region. (Alternatively: no member of the relevant series bears such a relation.) R bears the sufficiently intimate sorts of causal relations *only to other small regions*, in my view. So no material object that is wholly present at R is also wholly present at any large, human-shaped spacetime region.

[20] The endurantist solution to type C puzzles avoids coincidence (or co-location) by embracing multi-location. It might be argued, however, that multi-location is no less puzzling than co-location, at least when we're talking about *spatial* multi-location, in which a single thing is wholly present at two different places at one time, or in two different instantaneous spacetime regions that belong to a single ''simultaneity-slice'' or spacelike hypersurface. Since I have defended multi-location in detail elsewhere (2003), I will not address the above objection at length here. I would, however, like to make two brief points. (1) It seems to me that we can account for the initial plausibility of the ban on spatial multi-location, while allowing for the possibility of exceptions to this ban, in the following manner. First, we note the plausibility of the principle that a single material



I shall now present a second, physically more plausible, example of a type C situation. The General Theory of Relativity (GTR) permits the occurrence of what physicists call *closed timelike curves* (Gödel 1949), or *CTC*s. A timelike curve is a continuous path through spacetime corresponding to the possible life-history of a particle with mass. Unlike our time-traveling cell, a particle whose path is a timelike curve is ''always oriented towards [its] local future'':[21] at no point in its career does the object travel backward in time with respect to its immediate neighborhood in spacetime. A timelike curve is *closed* just in case it forms a loop, thus ''ending where it began'', so to speak. A particle that

object can be wholly present in two different spacetime regions only if the appropriate sort of immanent causal relation holds between the contents of those regions. Next, we note that only in the most bizarre circumstances could the relevant sort of causal relation hold between the contents of distinct but simultaneous spacetime regions. Hence, in the absence of such bizarre circumstances, spatially multi-located material objects are impossible, as many people normally assume. However, the time travel situations that I am now considering provide exactly the bizarre sorts of circumstances needed for the relevant type of causal relations to hold between simultaneous spacetime regions and hence for spatially multi-located material objects. (2) Objection: a spatially multi-located material object might have contrary properties at a single time. e.g. it might be that the whole of O is hot (over here) and, at the same time, the whole of O is cold (over there). But it's impossible for the whole of a thing to be both hot and cold at the same time. Reply: O bears the hot-at relation to spacetime region R (or to a moment pt of O's proper time at which O is wholly present at R) and O bears the cold-at relation to the different (but simultaneous) spacetime region R* (or to a moment pt* ($\neq$ pt) of O's proper time at which O is wholly present at R*). Alternatively, a similar relativizing treatment can be given, not to the variable properties (hotness and coldness), but the instantiation relation. For further defenses of multi-location, see McDaniel (2003) and Beebee and Rush (2003).

[21] In the words of Smith (1998).

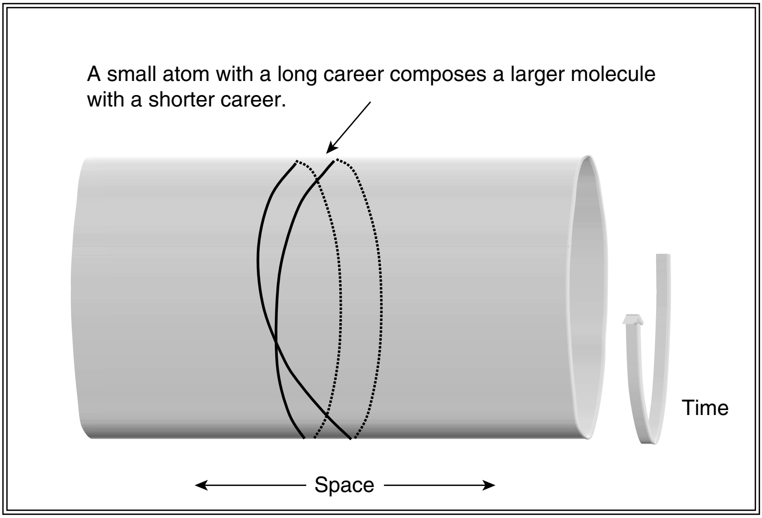


Figure 6.1 Time travel via a CTC in cylindrical spacetime

traces out an *almost* closed timelike curve would, just by lasting long enough and taking the appropriate trajectory, return to its own past and coexist with a younger version of itself.

Since the original discovery that GTR permits CTCs, physicists and philosophers have noted that similar time-travel scenarios can be constructed in the Minkowski spacetime of the Special Theory of Relativity and even in the pre-relativistic contexts of Newtonian and neo-Newtonian spacetime. We need only suppose that the relevant spacetimes could be 'cylindrical', with a closed, circular, temporal dimension.[22]

Consider the path in Figure 6.1, and suppose that it represents the career of a hydrogen atom, which we shall call 'Adam'. Adam is spatially bi-located throughout its two-billion-year-long career. For any given moment of external time (or 'global simultaneity slice') t in the relevant universe, Adam is present at t 'twice over': that is, there are *two* different moments pt and pt* of Adam's proper time such that, at pt, Adam is present at t, and at pt*, Adam is present at t. Suppose that, at each moment of Adam's proper time, Adam is chemically bonded to itself at a different moment of its proper time, thus forming a molecule of H_2, which we shall call 'Abel'. Abel is spatially mono-located

[22] See, e.g. Weingard (1979: 207), Horwich (1987), Earman (1995).

throughout its career (which is only one billion years long). For any given external time t, Abel is present at t only once: that is, there is only *one* moment of Abel's proper time at which Abel is present at t.

The case of Adam and Abel seems to be a type C situation. These objects trace out the same path over the course of their careers, but they are never spatially co-located. In particular, each of Adam's S-regions is smaller than any of Abel's S-regions. These objects share their path but none of their S-regions. The fact that they have different S-regions entails that they are distinct. But, as in the previous case, the distinctness of Adam and Abel can be argued for in a number of ways. Adam, being a mere hydrogen atom, has certain chemical properties that Abel lacks. Abel, being a hydrogen molecule, is more massive than Adam.[23]

This last point deserves elaboration. Let us think of a *mass history* as a certain sort of property of a material object, a property that reflects the way in which the object's mass changes, or stays constant, over the course of its career. Mass histories can be more or less specific. The following property, for example, is a fairly specific mass history:

> *being an object that has a rest mass of 10 units throughout the first ten years of its career, then gradually increases in mass for the next five years until it attains a rest mass of 20 units, at which point its rest mass remains constant for the final three years of its career.*

Here are two more mass histories, the first instantiated by Adam, the second by Abel.

MH1 *being an object that has a rest mass of 1 unit throughout its 2-billion-year-long career.*

MH2 *being an object that has a rest mass of more than one unit throughout its 1-billion-year-long career.*

Mass histories MH1 and MH2 are, of course, quite different: MH1 is characteristic of a hydrogen atom, whereas MH2 is characteristic of a molecule of H_2.

MH1 and MH2 seem to be incompatible intrinsic properties. Each seems to be *intrinsic* in the sense (roughly stated) that whether or not a

[23] We can imagine a variant of this case in which Adam is replaced by a small, long-lived sentient being who is elated throughout its life, and Abel is replaced by a larger, shorter-lived sentient being who is depressed throughout its life. The apparent incompatibility of these mental histories would seem to show that the creatures are distinct.

thing has the property depends only on what the thing is like in itself, and is independent of how that thing is related to anything else. They seem to be *incompatible* in the sense that it's impossible for a single thing to have both of them. Moreover, these properties seem to be purely physical, non-modal, and mind-independent. The fact that Adam has MH1 and Abel has MH2 would, therefore, seem to provide a very firm basis for the conclusion that Adam ≠ Abel. The force of this argument ought to be acknowledged by an extremely wide range of philosophers.

4. AN ARGUMENT AGAINST PERDURANTISM, AND SOME RESPONSES

Confronted with the puzzles that I have described, perdurantists have a number of options. To present these options, it will be useful to have my anti-perdurantist argument set out in numbered form:

P1 Type C situations are possible: there are possible worlds in which they occur.
P2 For any possible world w, if a type C situation occurs in w, then either: (i) there are numerically distinct material objects that coincide in w (in which case the anti-coincidence principle is false) or (ii) perdurantism is not true in w.
P3 The anti-coincidence principle is true: it is *impossible* for numerically distinct material objects to coincide.
C1 Perdurantism is not a necessary truth: there are possible worlds in which it is not true. (P1, P2, P3)
P4 Perdurantism, like endurantism, is either necessarily true or necessarily false.
C2 Perdurantism is necessarily false. (C1, P4)

As I noted at the beginning of the chapter, there are various ways in which the perdurantist can resist this argument. Each of these options, however, would significantly weaken at least one well-known anti-endurantist argument. Or so I shall now try to show.

First option. The perdurantist can deny P4 and hold, with David Lewis (1999: 227), that perdurantism is a contingent truth. Such a perdurantist could say that there are possible worlds at which type C situations occur and that perdurantism is false at such worlds. He could go on to say that

perdurantism remains true in the actual world, which does not (so far as we know!) contain any type C situations.

I am not entirely hostile to this response. First, it concedes most of what I have attempted to show in this chapter—namely, that there is a third and heretofore unnoticed type of coincidence situation, a type whose occurrence is compatible with endurantism but not with perdurantism (given the anti-coincidence principle).

Secondly, this response undermines all anti-endurantist arguments that rely on P4 (or something relevantly like it). One such argument is Theodore Sider's time-travel-based argument against endurantism (2001: 101–9). Sider tries to show that

(S) endurantism is false in worlds in which certain sorts of time travel occur.

But of course (S) by itself does not entail that

(T) endurantism is false in the *actual* world (where, so far as we know, the relevant sort of time travel does not occur).

As Sider is well aware, the move from (S) to (T) requires an additional assumption. Either (A1) or (A2) would do the necessary work:

(A1) a theory of persistence (such as endurantism or perdurantism) is false in the actual world if it is false in any possible world, or
(A2) it's true both that:

(i) a theory of persistence is false in the actual world if it is false in any possible world w that is governed by the same laws of nature as the actual world, and that
(ii) the relevant sorts of time travel are nomically possible: they occur at possible worlds that are governed by the same laws of nature as is the actual world.[24]

But any principle that could license Sider's inference from (S) to (T) would also, it seems, license the corresponding inference in my argument.[25] In other words, any such principle will be capable of playing the

[24] Gödel (1949) relies on a premise closely analogous to (A2) in his argument for the unreality of time. For insightful discussion of Gödel's argument, see Savitt (1994), Earman (1995), and Yourgrau (1999).

[25] Provided that it is stated as constraint on *any* theory of persistence, not merely as a special constraint upon endurantism. If it's stated in the latter fashion, then it will be plausible only to the extent that a corresponding constraint upon perdurantism is also plausible.

role of P4.[26] I conclude, therefore, that the perdurantist who blocks my argument by rejecting P4 (and its potential substitutes) thereby undermines Sider's time-travel-based argument against endurantism. While some of Sider's anti-endurantist allies might see this as a small price to pay, Sider himself rates the argument from time travel as one of the weightiest considerations against endurantism, ahead, for example, of Lewis's argument from temporary intrinsics (of which more later) and various arguments from relativity theory.

Second option. The perdurantist can deny P3 and hold that it's possible for two different material objects to coincide—that is, to exactly occupy the very same location and to be composed of the very same particles or matter at that location. This option, however, undermines all anti-endurantist arguments that depend upon the truth of P3, including Mark Heller's influential ''Body/Body-minus'' argument (1984). (Heller argues, roughly, that since the anti-coincidence principle is true, and since type A situations violate this principle given endurantism but not given perdurantism, the latter view must be correct.[27])

Third option. Perdurantists can deny the possibility of backward time travel[28] and with it the possibility of type C situations, which necessarily involve such time travel. This would allow perdurantists to reject P1. This third option has two strikes against it. First, it's implausible: the

[26] Consider (A1). Just as (A1) would allow Sider to derive the actual falsehood of endurantism from its possible falsehood, this assumption would allow me to derive the actual falsehood of perdurantism from its possible falsehood. Next consider (A2). It says (i) that if a theory of persistence is false at some nomically possible world, then it's false at the actual world and (ii) that the sorts of time travel situations that Sider takes to be incompatible with endurantism occur at nomically possible worlds. This would of course allow Sider to derive the actual falsehood of endurantism from its falsehood at the relevant time travel worlds; and it would allow me to derive the actual falsehood of perdurantism from its falsehood at the given worlds.

[27] Heller and his allies might reply by claiming that his argument relies not on the anti-coincidence principle (according to which coincidence is *impossible*) but on the weaker principle that distinct material objects do not *actually* coincide. Even this weaker principle (the reply continues) is enough to show that endurantism is false given something that many will grant—namely, the *actual* occurrence of a type A situation. In response, I shall content myself with the following point. It is unclear what motivation one could possibly have for holding the weaker principle but not the stronger. If one concedes that distinct material objects *can* coincide, and even that they do so in worlds governed by the same laws of nature as our own, why think that they don't *actually* coincide? This point seems sufficient to establish that, the above reply notwithstanding, the perdurantist who takes the ''second option'' thereby significantly weakens Heller's Body/Body-minus argument.

[28] Both via CTCs and via discontinuous jumps.

case in favor of the metaphysical possibility of backward time travel, even of the ''Wellsian'' variety involved in the Cell–Tubman example, is quite forceful.[29] An even more powerful case can be made for the ''Gödelian'' time travel involved in the Adam–Abel example, which seems to be not just metaphysically but also nomically possible.[30] Secondly, denying the possibility of backward time travel (like denying P4) would obviously undermine Sider's time-travel-based argument against endurantism.

Fourth option. Perdurantists can adopt an eliminativist view about composite material objects[31] and deny the possibility of composites like Tubman and Abel, composites that would—if they existed—have the same paths as the smaller, longer-lived objects that seem to compose them.[32] Like the third option, this would allow perdurantists to deny P1: whenever we may *seem* to have a larger object and a smaller object

[29] See, e.g. Lewis (1976) and Sider (2002).

[30] See, e.g. Horwich (1975), Weingard (1979), Earman (1995), and Smith (1998). The terms 'Wellsian' and 'Gödelian' are due to Earman.

[31] For a non-dismissive discussion of such a view, see Rosen and Dorr (2002).

[32] As Kris McDaniel has noted, it might be argued that composites such as Abel and Tubman are on worse footing than composites generally, since it may seem that the former objects but not the latter objects would violate the widely accepted mereological principle ('weak supplementation') that if a thing has a proper part, then it has another proper part that shares no part with the first. It may seem that all of Abel's parts share parts with Adam. In my view, the best response to this problem is to hold that the fundamental parthood relation is neither the two-place relation 'x is part of y' nor the three-place relation 'x is part of y at time t' but rather is the four-place relation 'x, at moment t_x of its proper time, is part of y, at moment t_y of its proper time'. Traditional formal theories of the part–whole relation assume that parthood is the two-place relation. Those who hold that parthood is three-place and temporally relativized can, however, formulate various analogues of the familiar principles of traditional mereological systems. Thus, e.g. the transitivity principle of traditional theories can be restated as: (T*) for any objects x, y, and z and time t, if x is part of y at t and y is part of z at t, then x is part of z at t. (The traditional principle has other, less plausible, analogues as well.) Similarly, if we assume that parthood is four-place, we can formulate various analogues of the principles stated in terms of two- or three-place parthood. The strongest plausible analogue of (T*) is: (T**) for any object x and moment of its proper time t_x, any object y and moment of its proper time t_y, and any object z and moment of its proper time t_z, if x at t_x is part of y at t_y and y at t_y is part of z at t_z, then x at t_x is part of z at t_z. What, then, of weak supplementation? Here is the strongest four-place analogue of that principle that all of my cases respect: (S**) for any object x and moment of its proper time t_x and any object y and moment of its proper time t_y, if [x at t_x is part of y at t_y and either ($x \neq y$ or $t_x \neq t_y$)], then there is some object z and moment of its proper time t_z such that: (i) z at t_z is part of y at t_y, (ii) either $z \neq y$ or $t_z \neq t_y$, (iii) either $z \neq x$ or $t_z \neq t_x$, and (iv) it's not the case that there's some object u and moment t_u of its proper time such that (a) u at t_u is part of x at t_x and (b) u at t_u is part of z at t_z. (The analogy may be easier to see if one notes that ordered

involved in a type C situation, we *in fact* have only the smaller object (if we have even that much).

Also like the third option, this fourth option has two strikes against it. First—and for what it's worth—this option is at least somewhat counterintuitive: when we have many particles arranged 'human-wise', it does intuitively seem that these particles compose a human being (or at least a humanoid); and when we have quarks and electrons arranged 'hydrogen molecule-wise', it intuitively seems that these particles compose a hydrogen molecule. The second strike against option four, from the point of view of perdurantism, is that it would significantly weaken Heller's Body/Body-minus argument against endurantism, which assumes that this sort of eliminativism is unacceptable as a solution to type A puzzles. (Sider's argument from vagueness (1997, 2001, forthcoming) against endurantism also conflicts with this eliminativist position.)

Fifth option. Perdurantists can hold that, despite the apparent differences between them, Adam and Abel are in fact numerically one and the same thing, and so too for Cell and Tubman. Like the two previous options, this fifth option would allow the perdurantist to reject P1 and deny the possibility of genuine type C situations (which by definition involve numerically distinct objects).

I argued, recall, that Adam and Abel have incompatible mass histories—that Adam has MH1 whereas Abel has MH2. Since it's impossible for a single thing to have incompatible properties, I concluded that Adam and Abel were distinct. Perdurantists, however, might respond as follows.

The thing that we're calling 'Adam' can be exhaustively partitioned into instantaneous temporal parts in different ways. On one way of being partitioned—call it the 'atomish' way—each of the relevant parts has the size, shape, and mass of a hydrogen atom. On a different way of being partitioned into instantaneous temporal parts—call it the 'moleculean' way—each of the relevant parts has the size, shape, and mass of a molecule of H_2. All of this is true of the thing we're calling

<object, moment of proper time> pairs play roles in the new principles that are similar to the roles played by objects in the traditional principles.) Although my cases do violate stronger versions of this principle, I do not find those versions plausible, given the availability of (S**). On weak supplementation in terms of two- and three-place parthood, see Simons (1987). For a consideration of the view that *coexistence* is four-place and relativized to moments of proper time, see Balashov (2000: 162 n. 15). In my 2003 and 2004 I show that treating apparently n-adic *spatial* relations as really being 2n-adic (and relativized in the relevant manner either to regions or moments of proper time) solves certain problems for immanent universals and enduring time-travelers.

'Abel' as well. But given these observations, it becomes plausible to say that Adam has mass history MH1 only relative to the atomish way of being partitioned into instantaneous temporal parts,[33] and that Abel has the *contrary* mass history MH2 only relative to a *different* way of being partitioned into instantaneous temporal parts.[34] And since it's possible for a single thing to have one mass history relative to one way of being partitioned and a contrary mass history relative to a different way of being partitioned, none of the facts about Adam's and Abel's mass histories entails that Adam ≠ Abel. Parallel maneuvers could be expected to block the other arguments for the distinctness of Adam and Abel (and of Cell and Tubman).

The perdurantist who adopts this 'relativizing' treatment of mass histories seems to be committed to some view along the following lines.[35] (1) Although mass histories may appear to be monadic, intrinsic properties of things, they are in fact disguised, dyadic relations that things bear to *ways of being partitioned into instantaneous temporal parts* (for short: to *partitions*). Thus, for example, rather than saying that Adam has property MH1, we should say that Adam bears the *MH1-relative-to* relation to a certain partition. (2) The dyadic having (having *simpliciter*, just-plain-having) of mass histories must be replaced by the triadic relation *having-relative-to*. Thus, rather than saying that Adam just plain *has* MH1, we should say that Adam bears the *having* relation to MH1 and a certain partition.

[33] What is a *way of being partitioned into instantaneous temporal parts*? I shall leave this question to perdurantists.

[34] This view might be combined with a view of mass history predicates that Noonan (1993) would call *Abelardian*. According to such a view, the property expressed by that predicate varies depending upon the sense of the name to which the predicate is attached. Thus, e.g. when attached to the name "Adam", the predicate "has mass history MH1" expresses the property of having mass history MH1 relative to the atomish partition, and when attached to the name "Abel", that same predicate expresses a different property—namely, that of having mass history MH2 relative to the moleculean partition. This would allow the perdurantist to concede that "Adam has MH1" is true while "Abel has MH1" is false without being forced to conclude that Adam ≠ Abel.

[35] To keep the discussion manageable, I will not attempt a comprehensive survey of relativizing responses to my argument for the distinctness of Adam and Abel. However, I do hope to say enough to make it clear that all of the likely relativizing responses to my argument will ultimately have the same effect—namely, that of undermining Lewis's argument from temporary intrinsics against endurantism. (For surveys of relativizing responses to Lewis's argument, see Lewis (2002) and Haslanger (2003). A number of these responses suggest corresponding responses to my argument for the distinctness of Adam and Abel.)

I concede that if the perdurantist applies some relativizing treatment of this sort to mass histories or the having of them, then he can resist my argument for the distinctness of Adam and Abel. It seems to me, however, that *if* the perdurantist makes this move, he will no longer be in a position to fault the endurantist who makes a parallel relativizing move in response to David Lewis's 'problem of temporary intrinsics'.[36] Here is Lewis:

Sometimes you sit, and then you are bent; sometimes you stand or lie, and then you are straight. How can one and the same thing have two contrary intrinsic properties? . . . I favor the hypothesis of perdurance. It says that persisting things are sums of temporal parts; their temporary intrinsic properties belong in the first instance to their temporal parts; and it is no problem that two different temporal parts can differ in their intrinsic properties. (2002: 1)

[The endurance solution, on the other hand, runs as follows.] Contrary to what we might think, shapes are not genuine intrinsic properties. They are disguised relations, which an enduring thing may bear to times. One and the same enduring thing may bear the bent-shape relation to some times, and the straight-shape relation to others. In itself, considered apart from its relations to other things, it has no shape at all. . . . This is simply incredible . . . If we know what shape is, we know that it is a property, not a relation. (1986: 204)

[An alternative endurantist solution holds that] it is not the intrinsic property *bent* or *straight*, but rather the copula that relates this property to a thing that has it, that turns into a relation to times. *Having* was originally thought to be a dyadic relation of things to properties; now it will instead be a triadic relation of things to properties and times . . . I protest that there is still nothing in the picture that has *bent* or *straight simpliciter*. (2002: 4–5)

A solution to the problem of intrinsic change for enduring things should . . . not replace monadic intrinsic properties by relations . . . [and] should not replace the having *simpliciter* of properties by standing in some relation to them. (2002: 1)

It seems to me that Lewis's criticisms of these relativizing maneuvers would apply with equal force to the relativizing maneuvers that the perdurantist needs to make in order to resist the argument for the distinctness of Adam and Abel. Let me explain.

Lewis's argument from temporary intrinsics against endurantism depends upon the following principle:

[36] I am grateful to Kris McDaniel for helping to clarify my thoughts on this point.

(L) For any material object O, if O changes from being bent to being straight, then:

(i) there is a thing that just plain *has* the monadic, intrinsic, non-indexed property *being bent*, and
(ii) there is a thing that just plain *has* the monadic, intrinsic, non-indexed property *being straight*, and
(iii) necessarily: for any x and y, if x just plain has the monadic, intrinsic, non-indexed property *being bent*, and y just plain has the monadic intrinsic, non-indexed property *being straight*, then x $\neq$ y.

From (L) it follows that whenever we have a thing O that changes from being bent to being straight, we have distinct things—proper temporal parts of O, presumably—one of which is bent, the other of which is straight. Endurantists reject (L). They can do so *apologetically*, conceding that (L) has some prima facie plausibility and that to reject it is to incur some cost, or *unapologetically*, arguing that (L) can be rejected at no cost at all.

Now, my main claim here is that if the endurantist incurs some cost in rejecting (L), then the perdurantist would incur some cost if he were to reject the parallel principles that can be used to argue for the distinctness of objects involved in apparent type C situations. One such principle is:

(L*) If a small hydrogen atom with a 2-billion-year-long career and a constant rest mass of 1 unit completely composes a larger hydrogen molecule with a 1-billion-year-long career and constant rest mass of more than 1 unit (in the manner illustrated by my case), then:

(i) there is a thing that just plain *has* the monadic, intrinsic, non-indexed property *being an object that has a rest mass of 1 unit throughout its 2-billion-year-long career*, and
(ii) there is a thing that just plain *has* the monadic, intrinsic, non-indexed property *being an object that has a rest mass of more than 1 unit throughout its 1-billion-year-long career*, and
(iii) necessarily: for any x and y, if x just plain has the monadic, intrinsic non-indexed property *being an object that has a rest mass of 1 unit throughout its 2-billion-year-long career* and y just plain has the monadic, intrinsic, non-indexed property *being an object that has a rest mass of more than 1 unit throughout its 1-billion-year-long career*, then x$\neq$y.

From (L*) it follows that in my Adam–Abel case, we have distinct things, one of which has a mass of 1 unit, the other of which has a mass of more than one unit. Presumably these things are Adam and Abel, respectively. The perdurantist who pursues option five and insists on identifying Adam with Abel must deny (L*).

I do not claim that this move is absolutely untenable. Perhaps it is, perhaps not. I claim only that, given the obvious parallels between principles like (L) and those like (L*), the perdurantist who rejects the latter incurs a cost comparable to that incurred by the endurantist who rejects the former. Thus the perdurantist who rejects (L*) and identifies Adam and Abel cannot in good conscience endorse Lewis's claim that the argument from temporary intrinsics is a ''decisive'' consideration against endurantism (1986: 203). So much, then, for the perdurantist's fifth option.

5. CONCLUSION

This survey of the perdurantist's options seems exhaustive: only by taking one or more of these options can the perdurantist resist my argument.[37] But, as I have noted, each of the options significantly weakens at least one important anti-endurantist argument. So, whether or not my new anti-perdurantist argument is sound, endurantism benefits.[38]

[37] Perhaps there is some way to solve type C puzzles by appeal to the sorts of nonstandard views about numerical identity (e.g. Geach 1980, Gallois 1998, T. Parsons 2000) that have been thought to solve puzzles of types A and B. Of course, the perdurantist who solves type C puzzles by appeal to such a view must abandon any argument against endurantism (e.g. Heller 1984) that assumes that such views are untenable as solutions to type A puzzles.

[38] Versions of this chapter were presented at Princeton University in the Spring semester of 2001, at the University of Nebraska at Omaha in March 2003, at the 2004 Central States Philosophy Conference in Iowa City, at the 2005 Central Division APA, and at the 2005 Bellingham Summer Philosophy Conference. Thanks to the members of those audiences, to the students in my 2004 and 2005 metaphysics courses at UNO, and to others with whom I've discussed the ideas in this chapter. They include John Carroll, Jerry Cederblom, Adam Elga, Gilbert Harman, Laura Grams, John Hawthorne, Mark Heller, Benj Hellie, Scott Jenkins, Halla Kim, Aaron Konopasky, Ned Markosian, Andrew Newman, Josh Parsons, Laurie Paul, Jim Pryor, Peter Vranas, and Dean Zimmerman. I am especially grateful to Yuri Balashov, Karen Bennett, Michael Glanzberg, Jeff Green, Mark Johnston, Simon Keller, Brian Kierland, Patrick X. Monaghan, Ted Sider, Nicholas J. J. Smith, Ryan Wasserman, and Brian Weatherson, all of whom provided extremely helpful written comments. My deepest debt is to Kris McDaniel, whose many rounds of insightful suggestions and criticisms have improved the chapter greatly.

REFERENCES

Balashov, Y. (2000) ''Enduring and Perduring Objects in Minkowski Spacetime'', *Philosophical Studies* 99: 129–66.

Beebee, H. and Rush, M. (2003) ''Non-Paradoxical Multi-Location'', *Analysis* 63.4: 311–17.

Davidson, M. (2004) ''Critical Notice of Theodore Sider's *Four Dimensionalism*'', *Philosophical Books*, January 2004, 1–17.

Earman, J. (1995) *Bangs, Crunches, Whimpers, and Shrieks* (Oxford: Oxford University Press).

Fine, K. (2003) ''The Non-identity of a Thing and its Matter'', *Mind* 112: 195–234.

Gallois, A. (1998) *Occasions of Identity: The Metaphysics of Persistence, Change, and Sameness* (Oxford: Clarendon Press).

Geach, P. (1980) *Reference and Generality*, 3rd edn. (Ithaca: Cornell University Press).

Gibbard, A. (1975) ''Contingent Identity'', *Journal of Philosophical Logic* 4: 187–222.

Gilmore, C. (2003) ''In Defence of Spatially Related Universals'', *Australasian Journal of Philosophy* 81: 420–8.

——(2004) *Material Objects: Metaphysical Issues*, Ph.D. thesis, Princeton University.

——(2006) ''Where in the Relativistic World Are We?'', *Philosophical Perspectives 20: Metaphysics*: 299–336.

Gödel, K. (1949) ''A Remark about the Relationship between Relativity Theory and Idealistic Philosophy'', in P. A. Schilpp (ed.), *Albert Einstein: Philosopher-Scientist* (La Salle, IL: Open Court), 557–62.

Haslanger, S. (2003) ''Persistence Through Time'', in M. Loux and D. Zimmerman, (eds.), *The Oxford Handbook of Metaphysics* (Oxford: Oxford University Press), 315–54.

Hawley, K. (2001) *How Things Persist* (Oxford: Oxford University Press).

Heller, M. (1984) ''Temporal Parts of Four Dimensional Objects'', *Philosophical Studies* 46: 323–34.

——(1990) *The Ontology of Physical Objects: Four-Dimensional Hunks of Matter* (Cambridge: Cambridge University Press).

Horwich, P. (1975) ''Some Alleged Paradoxes of Time Travel'', *Journal of Philosophy* 72: 432–44.

——(1987) *Asymmetries in Time* (Cambridge, MA: MIT).

Hudson, H. (2001) *A Materialist Metaphysics of the Human Person* (Ithaca: Cornell University Press).

Johnston, M. (1983) *Particulars and Persistence*, Ph.D. thesis, Princeton University.
——(1992) ''Constitution is not Identity'', *Mind* 101: 89–105.
Lewis, D. (1971) ''Counterparts of Persons and Their Bodies'', *Journal of Philosophy* 67: 203–11.
——(1976) ''The Paradoxes of Time Travel'', *American Philosophical Quarterly* 13: 145–52.
——(1986) *On the Plurality of Worlds* (Oxford: Blackwell).
——(1999) *Papers in Metaphysics and Epistemology* (Cambridge: Cambridge University Press).
——(2002) ''Tensing the Copula'', *Mind* 111(441): 1–13.
McDaniel, K. (2003) ''No Paradox of Multi-Location'', *Analysis* 63: 309–11.
Markosian, N. (2004) ''Two Argument's from Sider's *Four Dimensionalism*'', *Philosophy and Phenomenological Research* 68(3): 665–73.
Noonan, H. (1993) ''Constitution Is Identity'', *Mind* 102: 133–46.
Parsons, J. (2000) ''Must a Four-dimensionalist Believe in Temporal Parts?'', *The Monist* 83: 399–418.
——(2007) ''Theories of Location '', *Oxford Studies in Metaphysics*, Vol. 3. (Oxford: Oxford University Press).
Parsons, T. (2000) *Indeterminate Identity* (Oxford: Clarendon Press).
Rosen, G. and Dorr, C. (2002) ''Composition as a Fiction'', in R. Gale (ed.), *The Blackwell Guide to Metaphysics* (Oxford: Blackwell), 151–74.
Savitt, S. F. (1994) ''The Replacement of Time'', *Australasian Journal of Philosophy* 72: 463–74.
Sider, T. (1997) ''Four-Dimensionalism'', *The Philosophical Review* 106: 197–231.
——(2001) *Four Dimensionalism: An Ontology of Persistence and Time* (Oxford: Oxford University Press).
——(2002) ''Time Travel, Coincidences and Counterfactuals'', *Philosophical Studies* 110: 115–38.
——(2004) ''Replies to Gallois, Hirsch, and Markosian'', *Philosophy and Phenomenological Research* 68(3): 674–87.
——(forthcoming) ''Temporal Parts'', in John Hawthorne, Theodore Sider, and Dean W. Zimmerman (eds.), *Contemporary Debates in Metaphysics* (Basil Blackwell).
Simons, P. (1987) *Parts: A Study in Ontology* (Oxford: Oxford University Press).
Smith, N. J. J. (1998) ''The Problems of Backward Time Travel'', *Endeavour* 22/4: 156–8.
Weingard, R. (1979) ''Some Philosophical Aspects of Black Holes'', *Synthese* 42: 191–219.
Yourgrau, P. (1999) *Gödel Meets Einstein: Time Travel in the Gödel Universe* (Chicago: Open Court).

Part IV

MEREOLOGY

This page intentionally left blank

7. Theories of Location*

Josh Parsons

1. INTRODUCTION

Metaphysicians of space and time are fond of talking about objects being present at, wholly present at, or existing at certain times, or occupying certain regions of space, or even regions of spacetime. Take, for example, this famous set of definitions due to Mark Johnston and David Lewis:

> Let us say that something *persists*, iff, somehow or other, it exists at various times; this is the neutral word. Something *perdures* iff it persists by having different temporal parts, or stages, at different times, though no one part of it is wholly present at more than one time; whereas it *endures* iff it persists by being wholly present at more than one time. (Lewis 1986: 202)

A great deal of debate has been conducted in this terminology: debates about whether anything does endure or perdure; about the ontology of temporal parts; about whether it makes sense to apply this kind of thinking to space, as well as to time (we can ask, for example, the analogous questions whether things are extended by being entended, or pertended); about whether it can be applied to spacetime, and if so, to relativistic spacetime. These debates have been fruitful, but cursed with a certain amount of imprecision. People sometimes talk past each other because they mean subtly different things by ''is present at'', ''occupied'', and the like, or because they are simply confused about what they do mean.

* Thanks for comments on different versions of this material go to Yuri Balashov, Andrew Cortens, Michael Glanzberg, Cody Gilmore, John Hawthorne, Hud Hudson, Shieva Kleinschmidt, Dave Liebesman, Daniel Nolan, Ted Sider, Gabriel Uzquiano, Robbie Williams, my former colleagues and students at UC Davis and the University of St Andrews, and participants at the Bellingham Summer Philosophy Conference 2005.

In this chapter, I aim to clear up this problem of imprecision. I put forward a logical framework for discussing metaphysical debates that are couched in terms of what I call *location relations*—''present at'', ''occupies'', ''exists at'', and so on. This framework consists of a formal system whose theorems might be imagined to be all the conceptual truths about location relations (or perhaps even just all the truths of which it is not in dispute whether they are conceptual truths). I can then formulate, for example, the thesis that everything perdures as a sentence of my formal system, and see, in a formally rigorous way, what its consequences are. I can also show that we have a free choice of primitive location relations, and prove that certain location relations are interdefinable.

I describe my formal system as a theory of location because it is a theory in the logical sense—it is a set of axioms and definitions in the language of first order predicate calculus. But it isn't a theory in the sense of a complete metaphysical account of location. My goal is to construct a minimal background within which such complete accounts can be discussed.

It might be useful to contrast what I am doing here with some similar projects. I am not the first person to state a theory of location. Casati and Varzi do so in their (1999). Their system however, is much stronger than mine, and is not intended to be neutral with respect to metaphysical disputes. Casati and Varzi also do not put their system to temporal or spatio-temporal uses. I discuss their theory of location further in Section 7.

Many other people have attempted to regiment the Johnston/Lewis definitions of perdurance and endurance, or to define concepts like ''temporal part''. A good example can be found in Sider (2001). Sider actually has two definitions of ''temporal part'', and the difference between them is important. Sider's first definition, in terms of time-indexed mereology, is designed to appeal to his opponents; his second definition, in terms of atemporal mereology, is designed to express the theory he actually believes. It is this latter definition that is close to what I am attempting. Sider's definitions, however, both fall short of the generality I aspire to by assuming that if objects have proper temporal parts at all, then they have instantaneous ones. In Section 5 I show how to define ''temporal part'' in a way that does not assume this.

Finally, Gilmore (2006) gives a theory of location, informally, and with a very different structure from mine. He also defines ''endurance''

in a very different way from the way I would (and, I think, different from the Johnston/Lewis definition quoted above). I briefly say why I disagree with his treatment of the topic in Section 5.

2. LOCATION

Before we can think about temporal parts, we need to clarify what is meant by ''exists at'', or as I will say ''is located at''. This relation is not conceptually confined to times. We can also think of objects existing in places, or even regions of spacetime. The formal structure I will describe in this section is the same in each case. Indeed, it will be easiest if we begin by thinking about the spatial case.

Let us say that I am *weakly located* in my office iff I am in my office in the weakest possible sense: iff my office is not completely free of me.[1] I should count as weakly located in my office when I am sitting at my desk, when I am reaching an arm out of the window, or when I am reaching an arm in the window from the street outside. Let us say that I am *entirely located* in my office iff I am in my office and I am not anywhere outside my office; that is, iff I am in my office and everywhere outside my office is completely free of me. I am entirely in my office when I am sitting at the desk, but not when I am reaching an arm in or out of my window. Let us say that I pervade, or am *pervasively located* in, any place none of which is free of me. I don't ever pervade my office, but I do pervade the region exactly occupied by my left big toe. Finally, I am *exactly located* anywhere that I am both entirely and pervasively located. My exact location is like my shadow in substantival space.

These concepts are interrelated in interesting, formally expressible ways. To see their interrelations we need a way of formally expressing them:

$x@_{\circ}r$ for ''x is weakly located at r''
$x@_{<}r$ for ''x is entirely located at r''
$x@_{>}r$ for ''x pervades r''
$x@r$ for ''x is exactly located at r''

[1] In these examples, I am treating ''my office'' as the name of a *place*, not of a material thing. Or, speaking more strictly, I am assuming that ''my office'' names a place where things might be located; I am neutral as to whether that place is identical to some material things. This type of issue is discussed further in Section 8.

And we also need a way of expressing the relationships between the regions at which things are located: the relationship of one region being a *subregion* of another (as for example, my office is a subregion of the university campus). The subregion relation is intended to be reflexive—each region counts as its own subregion. A subregion of *r* that is not *r* itself is called a *proper subregion* of *r*. Another important relation between regions is that of *overlap*, or having a common subregion (as Australasia and Polynesia overlap on New Zealand):

$r < s$ for "*r* is a subregion of *s*"
$r \ll s$ for "*r* is a proper subregion of *s*"
$r \circ s$ for "*r* overlaps *s*"
$r \wr s$ for "*r* is disjoint from (i.e. does not overlap) *s*"

The four location relations given above can be inter-defined. An elegant way of thinking about them is to start with the idea of exact location, and then define the others in the following way: an object is entirely located in every super-region of its exact location; it pervades every subregion of its exact location; and it is weakly located at every region overlapping its exact location. This is shown formally in the definitions below (and I have chosen the notation for my location relations to remind us of these definitions):

(1) $x@_{<}r \equiv_{df} (\exists s)(x@s \wedge s < r)$
(2) $x@_{>}r \equiv_{df} (\exists s)(x@s \wedge r < s)$
(3) $x@_{\circ}r \equiv_{df} (\exists s)(x@s \wedge r \circ s)$

It's important, however, that we don't have to count exact location as a primitive of our theory. We could instead start with the more everyday concept of weak location—the "in" of "I am in my office".[2]

The definitions in that case are those suggested by the glosses I gave when I introduced the terms:

[2] A number of people objected to this. Am I really in my office if most of me is in the corridor, and just the tip of my toe is over the threshold? Well, suppose I arrive in the department and rush up to my office and stand in the corridor in just that way. At the same time a student asks the department secretary whether I am in my office. The secretary is in a position to know where I am. What should the secretary say?

It would not be crazy for the secretary to say "yes"—that shows that it is not crazy to think that I am in my office when just a little bit of me is. Of course there are also cases where it would be misleading for the secretary to say "yes" even though I am weakly located in my office. There must be a lot of pragmatic contextual goings on here—my only claim here is that weak location is closer to the everyday sense of "in" than exact location.

(4) $x@_{<}r \equiv_{df} x@_{\circ}r \wedge (\forall s)(x@_{\circ}s \rightarrow r \circ s)$
(5) $x@_{>}r \equiv_{df} (\forall s)(r \circ s \rightarrow x@_{\circ}s)$
(6) $x@r \quad \equiv_{df} (\forall s)(r \circ s \leftrightarrow x@_{\circ}s)$

The difference between (4) and (5) is noteworthy. Why does (4) have that extra first conjunct? If it were gone, then the definition of $@_{<}$ would just be the converse of the definition of $@_{>}$, which would seem more elegant. The reason is that I am not assuming that everything is in space. Suppose that the number seven is not in space. Then $(\forall s)(7@_{\circ}s \rightarrow r \circ s)$ is true, because the antecedent of the conditional is always false. But, in the sense of ''entirely located'' that I am trying to capture, the number seven is not entirely located everywhere; we need the extra conjunct to ensure that our definition of entire location agrees. Definitions (5) and (6) don't suffer from this problem, because there the weak location predicate occurs in the consequent of the conditional. To put the point another way, the definitions (5) and (6) already guarantee that, if $x@_{>}r$ or $x@r$, then $x@_{\circ}r$, so they don't need it as an explicit extra conjunct.

There are two very general principles about location that are worth noting here; the principles of functionality, and of exactness:

(Functionality) $(x@r \wedge x@s) \rightarrow r = s$
(Exactness) $(\exists r)(x@_{\circ}r) \rightarrow (\exists r)(x@r)$

Functionality says that exact location is a function. Exactness says that everything that is anywhere has an exact location. Each of them can be made to follow from definitions. Functionality is a consequence of (6) (given some further premises about the behaviour of $\circ$), and exactness is a very obvious consequence of (3). Both sound extremely plausible, given what's meant by ''exact location''.

People sometimes have doubts about Exactness. One reason for this doubt concerns objects part of which are in space, and part of which are not in space. Suppose that numbers are not in space, but that there are nonetheless ''mixed'' objects that have both numbers and concrete objects as parts. The mereological fusion of myself with the number seven (if there is such a fusion) would be such a mixed object.[3] In my

[3] If you have trouble imagining that numbers enter into mereological relations at all, you might change the example of a mixed object, to, say, the fusion of the myself with the singleton set of the number seven. It's not too hard to think of sets as *parts* of their supersets. Of course the mixed object is still pretty weird, and you might doubt that there are such things. If there aren't, then the reason for doubting exactness that I am discussing is doomed to begin with.

view, this object is exactly located just where I am (this does not follow from what I have said so far, but it does follow from the additional metaphysical principles described in Section 4). You might instead be tempted to say that this object is not exactly located anywhere, and indeed, that it is not entirely located anywhere, though it is weakly located everywhere that I am located.

I think that this kind of reasoning rests on a mistake about what ''entirely located'' means. Remember that for the mixed object to be entirely located somewhere is for there to be *no other place where it is*. The temptation to reject Exactness here is a temptation to treat objects not in space as if there were a special place, ''not in space'' in which those objects are spatially located. But that would be a mistake: an object not in space is not spatially located at all. (Later on, I'll draw a distinction between ''entirely located'' and ''wholly located''. Mixed objects won't count as *wholly* located anywhere.)

Another reason for rejecting Exactness has to do with the possibility of objects that are more spatially fine-grained than space itself. Since this is a complicated topic, I discuss it in its own section, Section 3. People also sometimes object to Functionality on the grounds that an enduring object might be exactly located at multiple times, or at multiple regions of spacetime. That, objection, I think, rests on a mischaracterization of endurance, and I discuss it in Section 5.

So far I've been using spatial examples. The concepts I've been using can easily be extended to the temporal case. We can say of John Locke, for example, that he is weakly but not pervasively temporally located in the year 1704 (the year of his death); that he temporally pervades the year 1700; and that he is exactly temporally located in the interval between a certain instant in the year 1646 and a certain instant in the year 1704. For those who like spacetime, we can also think of Locke having an exact location in spacetime, which would overlap both Oxford and the year 1700.

Let me also make some comments about my usage of ''time'', ''place'', ''region'', and so on. I mean all of these very broadly. I'm not assuming that there are points or instants, but I do mean ''time'' and ''place'' in a sense according to which if there are points, then they are places; and if there are instants, then they are times. I also count intervals, and any scattered times, as times, and I count both connected and scattered places as places, again if there are such things. I don't assume that any particular place or time is divisible or that any

particular set of places or times composes another (and for this reason I don't assume classical mereology); but I do stipulate that "is a place" and "is a time" are closed under division and composition. This is just what I mean by "place" and "time", not a deep metaphysical assumption.

3. EXACTNESS IN NON-STANDARD MODELS OF SPACE

One way that Exactness might be imagined to fail is if there could be objects that were more spatially fine-grained than space itself. Suppose that space is gunky—that every region has a proper subregion—and that every region of space has a volume greater than zero. There are, in short, no points of space. But it could still be (it seems) that there are point-shaped things. A point-shaped thing would be entirely located in each of a series of spherical regions of arbitrarily small volume, but would not pervade any such region. No point-shaped thing can be exactly located in or pervade any non-point-like regions. So point-shaped things would be entirely (and this weakly) located in a number of regions, but not exactly located anywhere. Though such a theory of space might seem odd, it seems coherent, and this (says the objector) shows that Exactness is not a conceptual truth of location.[4]

An application of this kind of conception of space appears in Daniel Nolan's (2006) work on the Stoic conception of mixture. Chrysippus, apparently, believed that when wine and water are mixed, none of the wine is transformed into anything other than wine, and nor is any of the water transformed into anything other than water. Moreover, nor is a mixture of wine and water like a mixture of wheat and lentils, with small blobs of pure wine rubbing shoulders with small blobs of pure water (Nolan calls this a "juxtaposition"). Rather, the wine–water mixture is "mixed through and through" so that no portion of the mixture is pure wine, and no portion is pure water.

There has been much debate over whether this idea of "mixture through and through" or "blending" is coherent. The trouble is to say what has happened to the portions of water and wine that were mixed. Have they been destroyed or transformed into something other than

[4] This type of objection to Exactness, involving point-shaped objects in gunky space, was given to me by Cody Gilmore; the following interpretation of Nolan, on which he denies Exactness, was suggested by Shieva Kleinschmidt.

pure water or pure wine? If so, then this is transformation, not blending. Are they still to be found, perhaps scattered, but untransformed, in the mixture? If so, then the water–wine mixture is a mere juxtaposition—not really any different from a mixture of wheat and lentils.

Nolan offers a modern reconstruction that aims to save blending from incoherence, and I here reconstruct his reconstruction in terms of the theory of location. The Nolanian Stoic believes in a gunky space in which every region is continuous, and that water and wine are themselves made out of homogeneous gunky matter. A blend of water and wine is such that every subregion of the region in which the blend is exactly located contains some water and some wine (unlike a mixture of wheat and lentils, some subregions of the exact location of which contain nothing but lentil or nothing but wheat). That is to say, after the wine and water are mixed, the wine is entirely located in the exact location of the blend, and weakly located in every subregion of the exact location of the blend, but does not pervade any such region. It follows from this that after blending, the wine is weakly located in all the regions that the blend is weakly located in, but has no exact location. The wine is so scattered and discontinuous that it will not exactly fit any of the continuous regions of Nolanian space.

I have my doubts whether this is really distinct enough from the wheat and lentils case to be what the Stoics were after. After all, on Nolan's view, there are still parts of the blend that are pure wine and pure water, rubbing shoulders in the blend just like grains of wheat and lentils. Unlike grains of wheat, they are not *exactly* located anywhere—but they are still parts of the blend, and they are still unmixed. Such an arcane metaphysical difference between blending and juxtaposition does not seem to do justice to the intuitive distinction between perfectly blended stuffs and jumbled piles of atoms. On the other hand, it may be that there is nothing better to answer to that distinction. Such an adept metaphysician and logician as Chrysippus would hardly believe in something that his contemporaries could easily tell was incoherent: that the blend has no part which is pure water, but yet the original pure water is part of it and has not changed its nature. Chrysippus *has to* agree that some parts of a blend are still pure, if he is not to be obviously contradicting himself; his view must therefore be that the pure parts of a blend are differently arranged from the pure parts of a juxtaposition. In any case, if Nolanian Stoic blends are coherent, Exactness is not a conceptual truth.

There's another way of generating counter-examples to Exactness in the same kind of spirit as those given above, but not involving gunky space. Someone might believe that space is so big that there is no maximal region. Let us call this conception of space, to coin a phrase, ''knuggy space''. Where the distinctive feature of gunky space is that every region of space has a proper subregion, the distinctive feature of knuggy space is that every region has a proper super-region. For all I've said in this chapter, space might be gunky, knuggy, both, or neither.[5] Just as no object can be exactly located at a point in gunky space (for there are none), no object can be exactly located at the maximal region in knuggy space (for there is none). If an object were located at the maximal region in a non-knuggy space, we might say that it was *omnipresent*. But, if you think about it, the theorist of knuggy space can (it seems) accommodate omnipresent objects, in a similar way to the way that the gunk theorist might accommodate point-shaped objects. An omnipresent object, we may say, pervades every region. If there is a maximal region, then it is entirely located there. If there is not, then it is not entirely located anywhere. In that latter case, we would have an object that is in space, but not exactly located anywhere.

These objections, I think, are the most serious problems with my framework so far. They may seem arcane, but I am making a very strong claim for Exactness—that it is a conceptual truth. It takes only the coherence of a counter-example to undermine that. What all three of these proposed counter-examples have in common is that they complicate the relationship between the geometrical properties of material things and the geometrical properties of space. They each also involve a counter-example to the schema ''x is S-shaped iff the exact location of x is S-shaped''. (Where S is ''point'', ''discontinuous'', and ''omnipresent'' respectively.) I think that there is something bad about that—though I find it hard to put my finger on. The following dilemma is my best attempt to draw out the problem.

A substantivalist, I think, would not much like these proposals because they require that we make sense of the geometrical properties of material things independently of the geometrical properties of places.

[5] One technical difference between gunky and knuggy space is that gunky space could still be a model of Classical Extensional Mereology (CEM), for CEM does not say that there must be atoms. Knuggy space, however, cannot, since CEM does require that there be a universe—a fusion of everything. Knuggy space is, however, a model of Minimal Extensional Mereology (MEM), the mereological system I am employing in this chapter.

To say ''there are things smaller than any region of space'', is to treat space like a kind of ghostly matter that happens to inter-penetrate all material things, but which might or might not have parts that match the sizes and shapes of material things. This immediately raises the question of why we should believe in such spooky stuff, since it seems objects could have all sorts of geometrical properties whether or not there are places around to be those objects' exact locations. Every substantivalist should be asked why, on their view, space is not just a kind of ghostly and immovable matter, and I do not know what the answer should be—but I think that allowing for the possibility of objects whose shapes do not exactly match the shapes of any region of space will make it harder to give one.

A relationalist about space, on the other hand, should also be dubious. Presumably the relationalist, at some point, is going to give a paraphrase of location talk into something relationalistically acceptable—some language not ontologically committed to regions of space. But why should the relationalist prefer to paraphrase one of these arcane, non-standard, gunky or knuggy theories of location when they could paraphrase a standard one instead? The non-standard models of space can only be attractive if they say something different from (and more correct than) the standard ones. So the relationalist can only be attracted to a non-standard model if she can say, in a relationalistically acceptable way, what it is getting right that the standard model is getting wrong. It is not at all clear what that could be.

4. PRINCIPLES OF PARTITION

The location predicates we've seen so far, and the principles of functionality and exactness have to do with how physical objects stand to places and times, and to parts of those places and times. Metaphysical doctrines about temporal parts, however, require us to talk about the parts of those physical objects, not just of places and times themselves.

In my presentation, the mereological relation does double duty both as the relation between times (and places) and their subregions, and as the relation between material things and their parts—between humans beings and feet, between computers and keyboards. This is for formal convenience more than anything else—it would be a hassle to have to define mereological notions such as overlap and fusion twice over, once

for times and places, and then again for material things. It's also fairly plausible as a metaphysical view that the places are *made out of* their subregions in the same sense that material things are made out of their parts. Should this seem objectionable, it is quite possible to remove this assumption while leaving intact everything else I say here. But it would be inconvenient, so I will leave the task to the reader who objects.

Using the formal apparatus of location relations and mereology, we can state some interesting and controversial metaphysical doctrines about parts and location. Consider, for example, the principle of arbitrary partition, which says that every object has a part exactly located in each region it pervades:

(Arbitrary partition) $x@_{>}r \rightarrow (\exists y)(y < x \wedge y@r)$

In the spatial case, this captures the idea that objects are divisible every way possible; every way corresponding to the ways space itself is divisible. There is such a thing (according to the principle) as the statue-shaped part of an uncut block of marble. For there is the region a statue would occupy, were one cut from the marble; and the marble pervades this region. So, by the principle, there is a part of the marble that is exactly located in that region. In the temporal case, this will turn out to be equivalent to the doctrine of perdurantism.

Arguments against this principle come in two varieties. There are arguments that purport to show that it leads to some absurd consequence, such as that no object can gain or lose parts (van Inwagen 1981). There are also arguments that purport to show that it is empirically false (Parsons 2000). I rather prefer the latter, but for the purposes of this chapter, I am going to confine myself to the claim that the principle is not a conceptual truth about location.

I have little to say by way of argument here: simply that I can conceive of an object being extended without having any proper parts. Such an object would pervade many regions without having parts exactly located at those regions. For example, there might be a completely solid sphere that had no proper parts. The left half of the region this sphere exactly occupies is pervaded, but not itself exactly occupied, by the sphere; and since the sphere has no proper parts, it has no proper part that exactly occupies that region either. Since I can conceive of such things, my concepts cannot be getting in the way; therefore, it is not a conceptual truth that every region pervaded by an object has a part of

that object exactly located in it. I'm not alone in this, either: there are a number of other philosophers who are in favour of extended simples.[6]

In thinking of such things, it's useful to make a distinction between being entirely located somewhere (in my terminology), and what is usually known as ''being wholly located'' somewhere. When I say that I am all in the office—that I am not stepping over the threshold or sticking an arm out of the window—there are two things I might mean. One is that I am *entirely* in the office, in the sense defined above: that everywhere disjoint from the office is free of me. The other thing I might mean is that I am *wholly* in the office: that none of me is missing from the office; every part of me is in the office.

It will be helpful to have a formal abbreviation for ''wholly located'':

$x@_{\blacktriangleleft}r$ for ''x is wholly located at r''

It can be defined in this way:

(7) $x@_{\blacktriangleleft}r \equiv_{df} (\forall y)(y < x \rightarrow y@_{\circ}r)$

Whole location and entire location tend to go around together. When I'm entirely in the office, I'm also wholly in the office; if I reach an arm out the window, I'm not entirely in the office, because I'm also in the street, but also, I'm not wholly in the office, because my arm isn't in the office. However, suppose the extended simple sphere were hovering over the sill of my window: it would be neither entirely in the office (for it is also in the street) nor entirely in the street (for it is also in the office). But it would be wholly in the office (for it has no part save itself, and that part is in the office) and wholly in the street (for it has no part save itself, and that part is in the street).

People sometimes give bad arguments for arbitrary partition by conflating ''wholly located'' with ''entirely located''. Of course, these labels I've chosen are technical—I don't claim that ''wholly'' and ''entirely'' have these different meanings in ordinary language. But it is one thing to say, ''I have no parts missing from here'', and another to say ''No place disjoint from here is a place where I am''. One statement

[6] Among them are, John Bigelow (1995: 21–7), Ned Markosian (1998), Fraser MacBride (1998: 220–7), Peter van Inwagen (1981), and, according to van Inwagen, Aristotle (van Inwagen 1990: 98). Spinoza (1994: 1P12, 1P13) claims that the universe itself is an extended simple. Democritus is usually understood to have held that the atoms come in an infinite variety of shapes and sizes, which would also commit him to extended simples.

asserts the non-existence of a certain kind of thing, the other the non-existence of a certain kind of place.

In the temporal case, the principle of arbitrary partition is the principle that objects have arbitrary temporal parts. People who believe the temporal version of the principle hold that persisting objects have proper parts exactly temporally located at each time they temporally pervade. For example, they hold that John Locke has a proper part that begins at the first instant of 1700, and ends at the last instant of that year. Those who are opposed to the temporal version of the principle hold that at least some things are wholly temporally located at some times at which they are not entirely temporally located. If Locke is such a thing, then he might be wholly temporally located in the year 1700, without being entirely temporally located in that year.

People who deny that persisting things have arbitrary temporal parts (or perhaps any proper temporal parts at all) are called *endurantists* and the things they believe in are said to *endure*, or to persist without having temporal parts—without *perduring*. Similarly, I say that some spatially extended things might have no arbitrary spatial parts (or perhaps no proper spatial parts at all). The things I am talking about, such as extended simples, may be said to be *entended*, or to be extended without having spatial parts—without being *pertended*.

Though I do not think that arbitrary partition principles are conceptually true, there are two related principles that seem as though they could be. Like arbitrary partition, they connect what sort of location an object has with what sorts of parts it has. The first of these is the principle of expansivity: the idea that an object cannot fail to be where its parts are. If part of me is in the office, then I can't be entirely absent from the office—my parts' exact locations must be subregions of my exact location.

There are a number of ways of stating this (another, weaker way is discussed in Section 7). Here is one that I like:[7]

(Expansivity) $x < y \wedge x@r \rightarrow (\exists s)(y@s \wedge r < s)$

This principle is respected by every ordinary example of location that I can think of. It is also helpful in thinking about exotic cases. Consider the example I mentioned earlier, in relation to the principle of exactness.

[7] The label ''expansivity'' that I use refers to Goodman's (1977: 38) idea of an expansive predicate—a predicate satisfied by anything that has a part satisfying that predicate.

Suppose that the number seven is not in space, and that there is a mereological fusion of me with the number seven. It follows from Expansivity that the fusion has an exact location which is a super-region of my exact location. Given that the fusion has an exact location at all, this seems right to me. However, it does not follow from Expansivity alone that the fusion's exact location is any particular super-region of my exact location.

It would follow from Arbitrary Partition and Expansivity that the fusion's exact location is identical to mine. Suppose, for *reductio*, that the fusion's exact location is a proper super-region of my exact location. Then there must be a subregion, *r*, of the fusion's exact location disjoint from my exact location, and pervaded by the fusion.[8] By Arbitrary Partition, then, the fusion has a part that is exactly located in *r*. But I am not located in *r* at all, so, by Expansivity, I do not overlap anything exactly located at *r*. Nor does the number seven, for the same reason. But the fusion of me and the number seven by definition overlaps only things that overlap either me or seven. We have a contradiction; so the fusion's exact location is not a proper super-region of my exact location. Expansivity requires that the fusion's exact location be a super-region, proper, or improper, of mine so they must be identical.

5. PERDURANCE, ENDURANCE, AND TEMPORAL PARTS

Consider again the Johnston/Lewis definitions of ''persists'', ''endures'', and ''perdures'':

> Let us say that something *persists*, iff, somehow or other, it exists at various times; this is the neutral word. Something *perdures* iff it persists by having different temporal parts, or stages, at different times, though no one part of it is wholly present at more than one time; whereas it *endures* iff it persists by being wholly present at more than one time. (Lewis 1986: 202)

The concepts of location are well geared to making these definitions more precise. Lewis's ''exists at'' is my ''weakly located in''; so an object persists iff it is temporally weakly located in many, disjoint, times. The proviso ''disjoint'' is needed here, because if an object exists for only an instant some time in 1973, then it is weakly located in 1973, and in the twentieth century, the second millennium and so on. That's not the kind

[8] This step employs the Strong Supplementation Principle discussed in Section 6.

of multiple location we wanted. This is not a counter-example to my claim that ''exists at'' means weak temporal location—most likely Lewis was tacitly restricting himself to instantaneous times in the quote above, which explains why he didn't need the proviso. It would certainly not help to identify ''exists at'' with exact temporal location, as no object is multiply exactly located.

Though temporal location has the same formal structure as spatial location, it is not the same relation. This may seem obvious, if you think that times and places are very different kinds of entity; but I don't wish to assume that. Imagine that someone other than John Locke happened to be born in the same instant, and die in the same instant as Locke himself. Locke and this other person, then, have the same temporal exact location, but not the same spatial exact location. To avoid confusion, then, I'll apply a superscripted T to location predicates, where I intend them to have a distinctively temporal meaning.

So, to give a formal version of Lewis's definition of ''persistence'':

(8) x persists $\equiv_{df} (\exists r)(\exists s)(r \wr s \wedge x@^T_\circ r \wedge x@^T_\circ s)$

There are two ways to formally define ''perdures''. One, which is suggested by the first clause of Lewis's definition, is in terms of the idea of a temporal part. This is slightly more complicated, so I will take the other route, and return to defining ''temporal part'' later. The second clause of Lewis's definition (and the only clause of his definition of ''endures'') suggests another way, in terms of ''wholly located'', which I defined in Section 4. A thing perdures iff it persists, and no part of it is wholly temporally located at two, disjoint, times. As before, we need the proviso ''disjoint'' because every object is wholly located both at its exact location, and all super-regions of its exact location:

(9) x perdures $\equiv_{df}$ x persists $\wedge (\forall y)(y < x \rightarrow (\forall r)(\forall s)$
$(x@^T_\blacktriangleleft r \wedge x@^T_\blacktriangleleft s \rightarrow ros))$

Now back to temporal parts. A perduring object has different temporal parts at different times. But just what is a temporal part? We actually have a three-place relation here—a relation between an object, a time, and its temporal part *at that time*. There are pitfalls that are often ignored in discussions of how to define this: What if we allow (as I do) times to include intervals? Or arbitrary fusions of intervals?

The best way to sort this out, I think, is to start with a spatial analogy. Though analogies between time and space are controversial, they are

less so among perdurantists, so using such an analogy to figure out the content of perdurantism should be unobjectionable.

When I reach my left arm out of the window of my office, there's a part of me that's in the street. Which part? My left arm. But of course if my left arm is in the street, then I have a lot of parts in the street: my left index finger, the fusion of all my left fingernails, and so on. When we say ''the part of me in the street'', we must mean something like ''the fusion of the parts of me in the street''. But that's not quite it, because I am a part of me, and I am in the street; so I, not my left arm, am the fusion of the parts of me in the street.

There are two ways to go here. We could say that my part in the street is the fusion of all of the parts of me that are wholly in the street; alternatively, we could say that my part in the street is the fusion of all of the parts of me that are entirely in the street.

To write the two definitions of temporal part down formally, we need a formal notation for mereological fusion. Following Simons (1987), I use a definite-description-like operator:

$(\sigma x)(\Phi x)$ for ''the fusion of everything Φ''

The two definitions of temporal part are given below. I am going to argue that it is the former that is the more useful notion; but since it will be helpful to have a label for the latter, I will call it ''temporal part*''. There is no issue of metaphysics (or even of conceptual analysis) at stake here. ''Temporal part'' is a technical term of metaphysics, and I'm simply defending my choice of stipulative definition.

(10) x's temporal part at $r =_{df} (\sigma y)(y < x \wedge y@^{T}_{\blacktriangleleft} r)$
(11) x's temporal part* at $r =_{df} (\sigma y)(y < x \wedge y@^{T}_{<} r)$

Unlike some definitions of temporal part, mine do not assume that time is made up of instants—cf. Sider (2001: 59–60). For all I've said, times might be gunky, every time having yet briefer times as parts, and never bottoming out in instants.

The difference between the two is how they behave when applied to enduring objects. If an object endures throughout its life, then its temporal part at each time it exists is the enduring object itself; this is because the object itself is wholly temporally located at each of the times it exists. But it has no part entirely temporally located at any subregion of the time at which it is exactly temporally located. So at some times at which it exists it has no temporal part*.

People sometimes think that temporal part* must be the concept at work when enduring objects are described as having ''no temporal parts''. That, however, is a mistake. It would be more consistent with the usual usage of mereological vocabulary to regard this as a typical shorthand for ''no temporal *proper* parts'', where a temporal proper part of an object is a temporal part (in my sense) which is also a proper part of that object. The notion of temporal part I prefer also makes it easier to give analyses of various time-indexed concepts that are neutral between enduring and perduring objects. And, as we will see, it makes it easier to make sense of the first clause of Lewis's definition of perdurance in terms of temporal parts.

Suppose we said that an object perdures iff it has different temporal parts at different times, as the first clause of Lewis's definition suggests. That definition of ''perdures'' would not be quite equivalent to the one in terms of ''wholly located''. They come apart, however, only in fairly exotic cases. Suppose that there's something that's made up of two parts, one of which perdures in the classical way, being wholly located at no time longer than an instant, the other of which has no proper parts at all. Suppose, for example, that immanent universals, such as *humanity*, have no proper parts; but that persons perdure in the classical way. Suppose that John Locke is the only instance of *humanity*; and call the fusion of Locke with *humanity*, Locke+. Locke+ is the sort of exotic object that makes our two definitions of perdurance come apart. Locke+ has a part which is wholly located at more than one disjoint time (namely, *humanity*), so Locke+ endures by the definition in terms of ''wholly located''. But Locke+ also has distinct temporal parts at each time, because every temporal part of Locke+ overlaps a different temporal part of Locke, so Locke+ perdures by the definition in terms of temporal parts.

It wouldn't help with this problem to replace ''temporal part'' in our definition with ''temporal part*''. That still makes Locke+ a perdurer by one definition, and an endurer by the other. Locke+'s temporal part* at each time is identical with Locke's temporal part at that time (except for times at which Locke+ is entirely located), so Locke+ again comes out having distinct temporal parts* at each time.

This shows that it would be better to understand ''different'' as it occurs in both places in Lewis's first clause as ''disjoint''. That will bring the two definitions into line, for though Locke+ has distinct temporal parts at each time, those temporal parts are not disjoint: they overlap on

humanity—and this is why Locke+ does not perdure. So our Lewis-inspired definition of perdurance in terms of temporal parts should read:

(12) x perdures $\equiv_{df}$ x persists$\wedge$
$(\forall r)(\forall s)(x@^{T}_{\circ} r \wedge x@^{T}_{\circ} s \wedge r \wr s \rightarrow x$'s temporal part at $r \wr x$'s temporal part at $s)$

An enduring object, according to Lewis, is one that is wholly located at more than one (disjoint) time. That is, any persisting object that does not perdure endures:

(13) x endures $\equiv_{df}$ $(\exists r)(\exists s)(r \wr s \wedge x@^{T}{}_{\blacktriangleleft} r \wedge x@^{T}{}_{\blacktriangleleft} s)$

This means that exotic objects like Locke+ would count as enduring. Also, objects like Johnston's (1987: 123–5) partial endurers—which mostly persist by enduring, but survive substantial changes only by perduring—would count as enduring by these standards. It would be useful to have a stronger notion of endurance that would require that an enduring object be wholly located at *every* time at which it exists:[9]

(14) x endures throughout $\equiv_{df}$ x persists $\wedge$ $(\forall r)(x@^{T}_{\circ} r \rightarrow x@^{T}{}_{\blacktriangleleft} r)$

What we've seen here is that we can get to the allegedly technical (and sometimes, allegedly incomprehensible) concepts of temporal part, perdurance, and so on by ratcheting up from mereological relations, subregion relations among times, and the concept of exact temporal location (or alternatively, weak temporal location).

My definitions provide a good answer to those endurantists who claim that, though they can understand classical mereology, they cannot understand what a temporal part would be. Moreover, since my definitions, and the theory of location as a whole, do not decide the question of endurantism versus perdurantism—and rightly so, I think, for it is not a question to be decided on conceptual grounds alone—those endurantists cannot claim that the question has been begged against them.

Some self-ascribing endurantists, however, have claimed that my theory of location begs the question against them. Moreover, though they might welcome the clarity of my terminology of location, they would also claim that I have mischaracterized endurantism. According to these people, an enduring object is not one that is wholly located at

[9] It still turns out that exotic objects that have some parts not in time at all do not endure throughout, even if those parts of them that are in time do.

each of many disjoint times, but one that is exactly located at each of many disjoint times. On this view, endurantism is the denial not of the temporal version of Arbitrary Partition, but of Functionality.

There are a number of things wrong with this rival interpretation of endurantism. First, there is my direct argument for Functionality: if someone doesn't believe Functionality is true, I begin to suspect they aren't talking about exact location—what part of ''exact'' do they not understand? There are a lot of other location relations they could be talking about instead; for example, the relation that holds between a thing x and a time t iff x is wholly and pervasively temporally located at t. That relation need not be a function, and indeed cases where it fails to be so are precisely cases in which x endures (in my sense).

Secondly, what unifies endurantists, on any interpretation, is their opposition to temporal parts. But the truth or falsity of Functionality has nothing to do with whether objects have temporal parts. If a thing could be exactly located at two disjoint times, t and t', it could do so and have a part exactly located at t but not t', and another part exactly located at t' and not t. On the proposed interpretation of endurance, the endurantist is not denying that objects have temporal parts. This is particularly bad because perdurantists, at least, will all agree with my characterization of *perdurantism*. If the endurantist has nothing to say that is incompatible with it, then there is no debate.

A good way of illustrating this, I think, is to refer to Cody Gilmore's (2006), which while otherwise excellent, embodies just the kind of confusion I am discussing here. Gilmore gives an informal theory of location, the location primitive of which he glosses both as ''wholly present'' and ''exactly occupies''. In his terminology, an ''enduring'' object ''exactly occupies'' many regions of spacetime, while a ''perduring'' object ''exactly occupies'' just one. This terminology could be brought into line with my own if ''exactly occupies'' were replaced by ''is wholly and pervasively located at''. But it is clear that Gilmore does not want that.[10]

It might seem tempting to interpret Gilmore's ''exactly occupies'' as my ''is exactly located at'', and treat him as a denier of Functionality. That, however, would be a mistake, for Gilmore has another relation between regions and objects which, though officially defined in terms of

[10] The reasons for this have to do with the details of Gilmore's argument. If Gilmore's ''exactly occupies'' were my ''is wholly and pervasively located at'', then his examples of Cell and Tubman would ''exactly occupy'' all and only the same regions of spacetime (whether or not they endure or perdure). It is crucial to Gilmore that this not be the case.

''is exactly located at'', is plainly my notion of exact location. Region *r* is the path of *x*, he says, iff *r* ''exactly contains *x*'s complete career or life-history''. That is what *I* mean by ''exact location'', and Gilmore does not deny that no object has more than one path, which is what it would take for him to deny Functionality. I think that this will turn out to be the case for anyone who seriously tries to pursue the view that enduring objects are multiply exactly located.

Gilmore's theory of location has the first problem I mentioned: it is more charitable to regard him as affirming Functionality than as denying it. It also has the second problem. Because Gilmore's ''exactly occupies'' is not explained in terms of location and mereology—and indeed cannot be—it would make the question of ''endurantism'' vs. ''perdurantism'' independent of the question of whether objects are divisible into arbitrary temporal parts. For all he has said, endurantism is compatible with everything advanced under the name ''perdurantism'' by Lewis, Sider, Quine, Smart, and so on. Though Gilmore's treatment of endurance perhaps characterizes what some (in my view, confused) endurantists would like to affirm, it mischaracterizes what they would like to deny.

6. THE FORMAL THEORY OF LOCATION

The time has come for some details. In this section I outline the formal nature of the theory of location, and contrast the two sets of definitions I gave in Section 2. The theory of location is built on a formal mereology. My mereological notation is as follows:[11]

$x \ll y$	for ''x is a proper part of y''
$x < y$	for ''x is a part, proper or improper, of y''
$x \circ y$	for ''x overlaps (i.e. has a part in common with) y''
$x \wr y$	for ''x is disjoint from (i.e. does not overlap) y''
$(\sigma x)(\Phi x)$	for ''the fusion of all that is Φ''

[11] I assume the classical first order predicate calculus with identity and definite descriptions. For clarity of notation, I use the convention that, where scope is left ambiguous, unary operators have narrower scope than binary operators, and conjunction and disjunction have narrower scope than implication. Also, where I use a sentence containing an unbound variable, that variable should be treated as if bound to a universal quantifier having the widest possible scope.

It's useful to distinguish a reflexive from an anti-reflexive sense of ''part''. Traditionally, the former is called ''part'', the latter ''proper part'', each thing being its own improper part. Either can be defined in terms of the other:

(15) $x \ll y \equiv_{df} x < y \wedge \neg x = y$
(16) $x < y \equiv_{df} x \ll y \vee x = y$

Overlap and disjointness can be defined so:

(17) $x \circ y \equiv_{df} (\exists z)(z < x \wedge z < y)$
(18) $x \wr y \equiv_{df} \neg(\exists z)(z < x \wedge z < y)$

A famous feature of formal mereologies is the notion of fusion. We may speak of the fusion of all the rabbits—intuitively, the thing that's made up of, and exhausted by, all the rabbits. This is usually written formally as a definite-description-like operator.[12]

(19) $(\sigma x)(\Phi x) =_{df} (\iota x)(\forall y)(x \circ y \leftrightarrow (\exists z)(\Phi z \wedge y \circ z))$

There's a common misapprehension about the concept of mereological fusion. People sometimes think of mereological fusions as being a distinctive class of entities, with, perhaps, their own distinctive range of essential properties. This way of thinking is perhaps suggested by an analogy with set theory, with which classical mereology is sometimes compared. Nothing could be further from the truth. Mereological fusion means nothing more nor less than what's said in the definition given above, and that definition is stated in a non-modal language. The mereological fusion of the Φs is just whatever it is that overlaps everything that overlaps a Φ. If that thing—whatever it is that overlaps everything that overlaps a Φ—is a bicycle, then it has whatever essential properties are appropriate to bicycles; if it is a hedgehog, then it has whatever essential properties are appropriate to hedgehogs. If it is something we have no name for apart from ''the mereological fusion of the Φs'', then for all we've said it has no essential properties at all. Someone who disagrees with this is simply speaking at cross-purposes with me: they have a stronger definition of ''fusion'' than that given by formal mereology.

Finally, for the mereological part of the theory of location, we need some mereological axioms. The following give us Simons's (1987: 25–31) *Minimal Extensional Mereology* (MEM):

[12] The definition of fusion given here is, strictly speaking, a definition schema.

(Asymmetry) $x \ll y \rightarrow \neg y \ll x$
(Transitivity) $x \ll y \wedge y \ll z \rightarrow x \ll z$
(Weak supplementation) $x \ll y \rightarrow (\exists z)(z \ll y \wedge z \wr x)$
(Maximal common part) $x \circ y \rightarrow (\exists z)(\forall w)$
$(w < z \leftrightarrow w < x \wedge w < y)$

There's another famous principle of mereology that's worth mentioning here, though it is *not* a theorem of MEM. This is the principle of unrestricted fusion. If it were added to the system, we would have *Classical Extensional Mereology* (CEM), which is strictly stronger than MEM:

(Unrestricted fusion) $(\exists x)(\Phi x) \rightarrow (\exists x)(\forall y)$
$(x \circ y \leftrightarrow (\exists z)(\Phi z \wedge y \circ z))$

To MEM we add the four location relations already described:

$x@_\circ r$ for "x is weakly located at r"
$x@_< r$ for "x is entirely located at r"
$x@_> r$ for "x pervades r"
$x@r$ for "x is exactly located at r"

There are two equivalent ways of defining these. We might make @ primitive, with functionality as an axiom. Call the system that includes MEM, the definitions (1), (2), and (3), and the functionality axiom, *S*@:

(1) $x@_< r \equiv_{df} (\exists s)(x@s \wedge s < r)$
(2) $x@_> r \equiv_{df} (\exists s)(x@s \wedge r < s)$
(3) $x@_\circ r \equiv_{df} (\exists s)(x@s \wedge r \circ s)$
(Functionality) $(x@r \wedge x@s) \rightarrow r = s$

The alternative is to have $@_\circ$ primitive, and exactness as an axiom. Call the system that includes MEM, the definitions (4), (5), and (6), and the exactness axiom, $S@_\circ$:

(4) $x@_< r \equiv_{df} x@_\circ r \wedge (\forall s)(x@_\circ s \rightarrow r \circ s)$
(5) $x@_> r \equiv_{df} (\forall s)(r \circ s \rightarrow x@_\circ s)$
(6) $x@r \equiv_{df} (\forall s)(r \circ s \leftrightarrow x@_\circ s)$
(Exactness) $(\exists r)(x@_\circ r) \rightarrow (\exists r)(x@r)$

Both of these systems affirm exactness and functionality. In *S*@, exactness is a consequence of (3). In $S@_\circ$, functionality is a consequence of (6) and the extensional overlap principle (20), which is itself a theorem of MEM:

(20) $x = y \leftrightarrow (\forall z)(x \circ z \leftrightarrow y \circ z)$

To see that exactness is independent of MEM plus the definitions of $S@_\circ$, consider the following counter-model: there are only three places, r, s, t, and one thing, a, all mereologically atomic. Object a is weakly located at r and s, but not at t. In this model, a has no exact location. This is not, of course, a model of CEM, which does not allow ''flat'' models like this one. Exactness would follow from (6) with the unrestricted fusion principle.

The definitions of $S@$ are also derivable in $S@_\circ$, in the sense that rules that would allow you to replace $@_\circ$, $@_<$ and $@_>$ with their definitions in $S@$ are derivable in $S@_\circ$. Similarly, the definitions of $S@_\circ$ are derivable in $S@$. In fact, the systems are equivalent.

7. CASATI AND VARZI

I'm not the first person to try to say something about the formal structure of location relations. It might help to contrast my proposals with the system described by Casati and Varzi (1999). Their theory of location is based on a formal mereo-topology they call GEMTC. For ease of presentation, I'll ignore the topological features of this theory, and describe it as it might be set up on the basis of a pure mereology. To this, Casati and Varzi add a primitive predicate of exact location. Like me, they treat the subregion relation among places and the part–whole relation among things in space as one mereological relation, and they define a number of inexact location relations:

(Partial Location) $PL(x,r) \equiv_{df} (\exists z)(z < x \wedge z@r)$
(Whole Location) $WL(x,r) \equiv_{df} (\exists s)(s < r \wedge x@s)$
(Generic Location) $GL(x,r) \equiv_{df} (\exists z)(\exists s)(z < x \wedge s < r \wedge z@s)$

Casati and Varzi's ''whole location'' is my entire location. Their ''partial location'' and ''generic location'' correspond to my pervasive and weak locations respectively (and this is made clear by the intuitive glosses they give on these concepts) but are defined differently.

They give the following axioms, intended to capture the formal features of location:

(Functionality) $(x@r \wedge x@s) \rightarrow r = s$
(Conditional reflexivity) $x@r \rightarrow r@r$
(Weak expansivity) $x < y \wedge x@r \wedge y@s \rightarrow r < s$
(Arbitrary partition) $r < s \wedge x@s \rightarrow PL(x,r)$

Functionality we have already discussed. The other three axioms each raise interesting questions.

Conditional reflexivity raises the question of what the domain of the location relations is. I've been loosely talking of ''things'' located in ''places''. But what about those places: where are they located, if anywhere? I do not see how to assign any metaphysical significance to this question. We might decide to say that every place is exactly located at itself (this is, effectively, what Casati and Varzi are saying here); or we might decide to say that no place is located anywhere (in the meaning of ''location'' in use). These seem to me to be equally good stipulations. So I have no quarrel with conditional reflexivity; but nor with its denial.

Casati and Varzi's expansivity principle is weaker than the one I described in Section 4. I describe the type of principle they use as ''weak expansivity''. Suppose that there are some objects that are outside space and time, which have no location at all. It is consistent with weak expansivity that these objects have parts that are in space or time; perhaps even all of their proper parts. Expansivity ought to rule this out, I think, and accordingly I recommend my principle as a friendly amendment to Casati and Varzi's system.

Finally, there is Casati and Varzi's version of the arbitrary partition principle, of which, as I've intimated, I take a dim view. I think that this axiom is an unwarranted metaphysical excrescence on a formal theory of location, and would be better removed. Unfortunately, it's not easy to remove it from Casati and Varzi's system, and the reasons why it is not easy shed some light on why people might mistakenly suppose that it is a conceptual truth. The reason has to do with the differences between Casati and Varzi's definitions of the inexact location predicates and my own.

Casati and Varzi's relation of Partial Location corresponds to my pervasive location; and Generic Location corresponds to my weak location. But the definitions are very different (and in the context of my system, inequivalent). To see this, begin by noting that *WL* and *PL* are not, as Casati and Varzi claim them to be, dual concepts. (Casati and Varzi 1999: 120) $WL(x, r)$ says there is a certain kind of place, namely x's exact location, and that r is related to it in a certain way. $PL(x, r)$, by contrast, says that there is a certain kind of *thing*, namely a part of x.

If Casati and Varzi really want the dual to *WL*, they should define it this way:

(21) $PL'(x,r) =_{df} (\exists s)(r < s \wedge x@s)$

This PL' relation is what I've called ''pervasive location''. Once we see the distinction between PL and PL', it's easy to see why Casati and Varzi want the arbitrary partition axiom. It erases that distinction, for in their system, it turns out, $PL'(x,r) \leftrightarrow PL(x,r)$ is a theorem, and Arbitrary Partition is needed to get the left-to-right conditional.

Without Arbitrary Partition, the definitions of PL and GL would be too strong to capture the intuitive meaning that Casati and Varzi's glosses make it clear that they are supposed to have. Suppose that I stand in the doorway of my office, one foot on the floor of the office, another foot on the floor of the corridor. I want to say that I am inexactly located in my office, in some sense. Here's a good sense in which I am: my exact location and my office overlap; they have an intersection—a region which is part of both. But, absent Arbitrary Partition, that doesn't suffice for my being inexactly in my office in any of Casati and Varzi's senses. They require that in addition to the intersection existing, there is some part of me that exactly occupies the intersection.

This requirement is gratuitous: if we used PL' instead of PL, and made similar revisions to the definition of GL, it would go away. The revised definitions would capture the intuitive meaning PL and GL are supposed to have without the need for Arbitrary Partition, rendering Arbitrary Partition (quite properly) an independent extra that might be affirmed or denied according to metaphysical taste. Besides that, the revised definitions would better capture the formal relationship between PL and WL. This is just what I've done in my definitions of ''pervades'' and ''entirely located''.

8. THINGS AND PLACES

The theory of location has a striking prima facie ontological commitment: it is substantivalist, and in this it reflects the ordinary way we talk about places and times. It also reflects the way philosophers of time have talked about persistence in terms of ''existing at'' multiple times. Though we talk this way, however, many of us are worried about the ontological commitment to places and times—are there really such things? If so, what are they like? If not, how can we understand talk that is apparently about them?

These questions are particularly pressing when we use the theory of location as a way of expressing theories about temporal parts. Participants in the temporal parts debate do not take themselves to be

exploring a number of variants of substantivalism; rather, they are asking questions that should be, for the most part independent of substantivalism.

There are three general answers to the problem posed by apparent commitment to substantivalism. First up, we could go for a form of *relationalism*: say that the formalism of the theory of location is metaphysically misleading, precisely because of its commitment to times and places. But then we have a problem—how do we understand discussions of temporal parts, of endurance, or ''wholly located''? Perhaps such talk could be paraphrased into a suitable relationalist framework. I have some hope of doing this myself, but it would be a difficult task, and beyond the scope of this chapter.

The second option I call *anti-reductionist substantivalism*. On this view, we take the seeming commitments of the theory of location with full metaphysical seriousness. There are fundamentally at least two kinds of object in the world, material ones (things) and immaterial ones (places); and there is a fundamental and unanalyzable external relation of location (exact or weak) between them. Nothing can be located at a thing; perhaps, with Casati and Varzi, we should say that each place is exactly located at itself; perhaps that no place can be located anywhere.

There might also be mixed objects—mereological fusions of things and places—some of whose parts are material and some immaterial. These seem like odd creatures. Where should we say that they are located? If each place is exactly located at itself, and strong expansivity is true, then a fusion of a thing x with a place r should at least pervade the fusion of r with the exact location of x. Perhaps we should say that mixed objects are always exactly located at the fusion of the exact locations of their parts. Though this answer is not arbitrary, it does not exactly help us understand what a mixed object might be like. Perhaps we should say that there cannot be mixed objects. That would be compatible with all I have said above; but it would not be compatible with a principle of unrestricted mereological fusion, which many people who would like to use the theory of location may be attracted to.

Here is another problem. Suppose a material object, x, is cubical. We are accustomed to thinking that shape properties are intrinsic. But surely, whether x is cubical depends on what sort of place it is exactly located in. A cubical object can't be located at a non-cubical place; but now we have a necessary connection between the intrinsic nature of

x and the external relations it bears to places, and indeed between the intrinsic nature of x and the intrinsic nature of its exact location. That should be impossible—it is a violation of a necessary condition on the set of intrinsic properties (Weatherson 2001: 369–73).

A good way to solve both these problems at one blow is to say that ordinary material objects, like chairs and tables, and you and me, *are* mixed objects; specifically, that we have our exact locations as proper parts. That way, *being cubical* is like having a nose—it is an intrinsic relational property[13]—and the necessary condition on the set of intrinsic properties is not violated because there is only a necessary connection between the intrinsic nature of x and the external relations it bears to some of its own parts—to things that are not wholly distinct from it.

This leaves us with another difficulty. What about the purely material part of an ordinary cube? On our current proposal, there must be such a thing; and it cannot be cubical in any ordinary sense. Where is it located? Perhaps it is not in space at all—that would be weird. But if it is in space, then where? Is it exactly located just where the ordinary cube is? If so, it would seem to exactly occupy a cubical place without itself being cubical.

The third possibility is *reductionist substantivalism*. I have taken care to ensure that the theory of location is compatible with a striking and surprising thesis: that each thing is exactly located at itself and nowhere else. This might be called the ''identity theory of location'':[14]

(Identity Theory of Location) $x@r \leftrightarrow x = r$

The typical metaphysical development of this view holds that places are much as the anti-reductionist substantivalist thinks they are—parts of a (mostly) immaterial plenum filling the entire universe—material

[13] Intrinsic properties are ones which ''concern how their instances are in themselves'' and can be contrasted with extrinsic properties. Relational properties are those that in some sense ''involve a relation''. In this terminology, due to Lloyd Humberstone (1996), relational properties may be intrinsic or extrinsic.

[14] John Hawthorne suggested an interesting variant on the identity theory of location to me—the ''coincidence theory'' of location. The coincidence theory would hold that x is exactly located at r iff x overlaps all and only the same things that r does. In the context of MEM, this is equivalent to the identity theory, but in a weaker, non-extensional mereology, it would not be. Since any two coinciding things must have all and only the same parts, we would still have Arbitrary Partition as a consequence. However, the coincidence theory could be held by a anti-reductionist substantivalist, since it allows for things to be distinct from their locations.

things, however, are boldly identified with ''matter filled'' places. ''Matter filled'', here, however, must not be understood to express a relation to some piece of matter that fills a place. Rather, it is an intrinsic property of some places that they are matter-filled. One feels rather as though this view is an eliminative reduction of matter, though its advocates might not see it that way.

The drawback of this is that the principle of arbitrary partition follows from the identity theory. (A proof of this is sketched in the appendix.) So, if you are not attracted to the principle of arbitrary partition, you will not want to be a reductionist substantivalist. Sider (2001: 110–19) has used this type of argument as an argument *for* arbitrary partition, at least in the temporal case. We should be substantivalists, he thinks, and substantivalists should be reductionists, or risk miring themselves in the sorts of difficulties described above. But reductionistic substantivalism entails arbitrary partition, so we should believe in arbitrary partition.[15] It would go beyond the scope of this chapter to pass judgment on this argument here, but I think that it is worthy of attention.

APPENDIX

In this section I present some theorems of the theory of location alluded to in the text. The proofs are given in a slightly abbreviated natural deduction form. They are intended as examples of how formally to reconstruct some of the arguments given in the text, rather than being of any logical interest.

Functionality is a theorem of $S@_\circ$ The functionality principle is an axiom of $S@$. Showing that it is a theorem of $S@_\circ$ is part of what's needed to show that the two systems are equivalent, as discussed in Section 6.

(22)	$x@r \wedge x@s$	(assumption for conditional proof)
(23)	$(\forall t)(x@_\circ t \leftrightarrow t \circ r)$	(22, $\wedge$-elim, definition of @)
(24)	$(\forall t)(x@_\circ t \leftrightarrow t \circ s)$	(22, $\wedge$-elim, definition of @)
(25)	$(\forall t)(t \circ r \leftrightarrow t \circ s)$	(23, 24)
(26)	$(\forall t)(t \circ r \leftrightarrow t \circ s) \rightarrow r = s$	(theorem of extensional mereology)

[15] The argument for substantivalism as against relationalism cited here (though not the argument for reductionist as against anti-reductionist substantivalism) is given in greater detail in Hawthorne and Sider (2002).

(27)	$r = s$	(25, 26, modus ponens)
(28)	$x@r \wedge x@s \rightarrow r = s$	(27, conditional proof, discharging 22)

Adequacy of *S@*-definitions in *S@*$_\circ$ The location predicates are defined differently, and in terms of a different primitive in $S@$ and $S@_\circ$. However, they are, in an important sense, synonymous. One way to show this formally is to show that the definitions of each system are derived rules of the other. It would be a long and boring task to do this for each definition. Instead I will give a couple of illustrative examples.

I start by deriving the definition of $@_\circ$ in terms of $@$ in $S@_\circ$. That is, I am going to show that $x@_\circ r$ and $(\exists s)(x@s \wedge r \circ s)$ are inter-derivable in $S@_\circ$. The other definitions of $S@$ can be derived in $S@_\circ$ in a similar way. First, the right-to-left direction, which is very easy:

(29)	$(\exists s)(x@s \wedge r \circ s)$	
(30)	$(\exists s)((\forall t)(x@_\circ t \leftrightarrow t \circ s) \wedge r \circ s)$	(29, definition of @)
(31)	$(\forall t)(x@_\circ t \leftrightarrow t \circ s) \wedge r \circ s$	(30, ∃-elimination)
(32)	$(\forall t)(x@_\circ t \leftrightarrow t \circ s)$	(31, ∧-elimination)
(33)	$x@_\circ r \leftrightarrow r \circ s$	(32, ∀-elimination)
(34)	$r \circ s$	(31, ∧- elimination)
(35)	$x@_\circ r$	(33, 34, modus ponens)

Now the left-to-right direction. To do this we need to appeal to the exactness axiom of $S@_\circ$:

(36)	$x@_\circ r$	
(37)	$(\exists s)(x@_\circ s) \rightarrow (\exists s)(\forall t)(x@_\circ t \leftrightarrow t \circ s)$	(exactness axiom)
(38)	$(\exists s)(x@_\circ s)$	(36, ∃-introduction)
(39)	$(\exists s)(\forall t)(x@_\circ t \leftrightarrow t \circ s)$	(37, 38, modus ponens)
(40)	$(\forall t)(x@_\circ t \leftrightarrow t \circ s)$	(39, ∃-elimination)
(41)	$x@_\circ r \leftrightarrow r \circ s$	(40, ∀-elimination)
(42)	$r \circ s$	(36, 41, modus ponens)
(43)	$(\forall t)(x@_\circ t \leftrightarrow t \circ s) \wedge r \circ s$	(40, 42, ∧-introduction)
(44)	$(\exists s)((\forall t)(x@_\circ t \leftrightarrow t \circ s) \wedge r \circ s)$	(43, ∃-introduction)
(45)	$(\exists s)(x@s \wedge r \circ s)$	(44, definition of @)

So, in $S@_\circ$:

(46) $x@_\circ r \dashv\vdash (\exists s)(x@s \wedge r \circ s)$

Adequacy of $S@_\circ$-definitions in $S@$ Similarly, it would be nice to derive the definitions of $S@_\circ$ in $S@$. My example will be the definition of @—I will show that $x@r$ and $(\forall s)(r \circ s \leftrightarrow x@_\circ s)$ are inter-derivable. This proof appeals to the functionality axiom of $S@$, and to some theorems of mereology. First, the left-to-right direction:

(47)	$x@r$	
(48)	$r \circ s$	(assumption)
(49)	$(\exists t)(x@t \wedge t \circ s)$	(47, 48, $\exists$-introduction)
(50)	$r \circ s \rightarrow (\exists t)(x@t \wedge t \circ s)$	(49, conditional proof, discharging 48)
(51)	$(\exists t)(x@t \wedge t \circ s)$	(assumption)
(52)	$x@t \wedge t \circ s$	(51, $\exists$- elimination)
(53)	$x@r \wedge x@t \rightarrow r = t$	(functionality axiom)
(54)	$r = t$	(52, 53, modus ponens)
(55)	$r \circ s$	(52, 54, $\wedge$-elimination, substitution)
(56)	$(\exists t)(x@t \wedge t \circ s) \rightarrow r \circ s$	(55, conditional proof, discharging 51)
(57)	$(\forall s)(r \circ s \leftrightarrow (\exists t)(x@t \wedge t \circ s))$	(50, 56, $\forall$-introduction, $\leftrightarrow$ *-introduction*)
(58)	$(\forall s)(r \circ s \leftrightarrow x@_\circ s)$	(57, definition of $@_\circ$)
(59)	$x@r \vdash (\forall s)(r \circ s \leftrightarrow x@_\circ s)$	

And now the right-to-left direction. This is more difficult, so I have annotated the proof in more detail. Our first task is to show that (60) entails that x has an exact location, and give that exact location a name, u:

(60)	$(\forall s)(r \circ s \leftrightarrow x@_\circ s)$	
(61)	$(\forall s)(r \circ s \leftrightarrow (\exists t)(x@t \wedge t \circ s))$	(60, definition of $@_\circ$)
(62)	$r \circ r \leftrightarrow (\exists t)(x@t \wedge t \circ r)$	(61, $\forall$-elimination)
(63)	$r \circ r$	(theorem of extensional mereology)
(64)	$(\exists t)(x@t \wedge t \circ r)$	(62, 63, modus ponens)
(65)	$x@u$	(64, $\exists$-elimination, $\wedge$-elimination)

Now we prove that u is identical to r, using the extensional overlap principle (20) mentioned in Section 6. First we show that if u overlaps an arbitrary s, then r does too:

(66)	$u \circ s$	(assumption)
(67)	$x@u \wedge u \circ s$	(65, 66)
(68)	$(\exists t)(x@t \wedge t \circ s)$	(67, ∃-introduction)
(69)	$r \circ s \leftrightarrow (\exists t)(x@t \wedge t \circ s)$	(61, ∀-elimination)
(70)	$r \circ s$	(68, 69, modus ponens)
(71)	$u \circ s \rightarrow r \circ s$	(70, conditional proof discharging 66)

Now we show that if r overlaps an arbitrary s, so does u:

(72)	$r \circ s$	(assumption)
(73)	$(\exists t)(x@t \wedge t \circ s)$	(72, 69, modus ponens)
(74)	$x@t \wedge t \circ s$	(73, ∃-elimination)
(75)	$x@u \wedge x@t \rightarrow u = t$	(functionality axiom)
(76)	$u = t$	(65, 74, 75)
(77)	$u \circ s$	(74, 76, substitution, ∧-elimination)
(78)	$r \circ s \rightarrow u \circ s$	(77, conditional proof discharging 72)

But then, r and u overlap all and only the same things, and so are identical, by the extensional overlap principle. So, since x is exactly located at u, x is exactly located at r:

(79)	$r \circ s \leftrightarrow u \circ s$	(71, 78)
(80)	$(\forall s)(r \circ s \leftrightarrow u \circ s)$	(79, ∀-introduction)
(81)	$r = u \leftrightarrow (\forall s)(r \circ s \leftrightarrow u \circ s)$	(theorem of extensional mereology)
(82)	$r = u$	(80, 81)
(83)	$x@r$	(65, 82)

So, in $S@$:

(84) $x@r \dashv\vdash (\forall s)(r \circ s \leftrightarrow x@_{\circ}s)$

The other definitions of $S@_{\circ}$ can be derived in a similar way.

Arbitrary partition is entailed by the Identity Theory of Location
In Section 8 I claimed that reductionistic substantivalism entails the arbitrary partition principle, via the ''Identity Theory of Location''—the view that each thing is exactly located only at itself.

(85)	$x@_{>}r$	(assumption for conditional proof)
(86)	$(\exists s)(x@s \wedge r < s)$	(85, definition of $@_{>}$)

(87)	$x@s \wedge r < s$	(86, $\exists$-elimination)
(88)	$x@r \leftrightarrow x = r$	(identity theory of location)
(89)	$x = s$	(87, 88, modus ponens)
(90)	$r < x$	(87, 89, $\wedge$-elimination, substitution)
(91)	$r = r$	(=-introduction)
(92)	$r@r$	(91, 88, modus ponens)
(93)	$(\exists y)(y < x \wedge y@r)$	(90, 92, $\exists$-introduction, $\wedge$-introduction)
(94)	$x@_{>}r \rightarrow (\exists y)(y < x \wedge y@r)$	(93, conditional proof, discharging 85)

REFERENCES

Bigelow, J. (1995) *The Reality of Numbers* (Oxford: Clarendon Press).

Casati, R. and Varzi, A. C. (1999) *Parts and Places* (Cambridge, MA: MIT Press).

Gilmore, C. (2006) ''Time Travel, Coinciding Objects, and Persistence'', *Oxford Studies in Metaphysics*, Vol. 3 (Oxford: Oxford University Press).

Goodman, N. (1977) *The Structure of Appearance*, 3rd edn., (Dordrecht: D. Reidel).

Hawthorne, J. and Sider, T. (2002) ''Locations'', *Philosophical Topics* 30: 53–76.

Humberstone, L. (1996) ''Intrinsic/extrinsic'', *Synthese* 108: 205–67.

Johnston, M. (1987) ''Is there a Problem about Persistence?'', *Aristotelean Society Supp*. Vol. 61: 107–35.

Lewis, D. (1986) *On the Plurality of Worlds* (Oxford: Blackwell).

MacBride, F. (1998) ''Where are Particulars and Universals?'', *Dialectica* 52(3): 203–27.

Markosian, N. (1998) ''Simples'', *Australasian Journal of Philosophy* 76(2): 213–28.

Nolan, D. (2006) ''Stoic Gunk'', *Phronesis* (forthcoming).

Parsons, J. (2000) ''Must a four-dimensionalist believe in temporal parts?'', *Monist* 83(3): 399–418.

Sider, T. (2001) *Four-dimensionalism* (Oxford: Oxford University Press).

Simons, P. (1987) *Parts: A Study in Ontology* (Oxford: Clarendon Press).

Spinoza, B. (1994) *Ethics*, tr. Edwin Curley (London: Penguin).

van Inwagen, P. (1981) ''The Doctrine of Arbitrary Undetached Parts'', *Pacific Philosophical Quarterly* 62: 123–37.

——(1990) *Material Beings* (Ithaca, NY: Cornell University Press).

Weatherson, B. (2001) ''Intrinsic Properties and Combinatorial Principles'', *Philosophy and Phenomenological Research* 63(2): 365–80.

8. Brutal Simples

Kris McDaniel

1. INTRODUCTION

An object is a simple if and only if it has no proper parts. (*x* is a proper part of *y* just in case *x* is a part of *y* but *x* is not identical to *y*.) This is a *definition* of the word ''simple'', not a substantive criterion for being a simple. The *Simple Question* asks ''under what circumstances is a material object a simple?''[1] An answer to the Simple Question is an *informative* instance of the following schema:

Necessarily, *x* is a simple if and only if __________.

In other words, an answer to the Simple Question must provide necessary and sufficient conditions for being a simple, and it must not employ a mereological term on the right-hand side of the biconditional. An answer to the Simple Question is a substantive criterion for being a simple.

I will argue that there is no correct, finitely statable, and non-circular answer to the Simple Question. There is no non-mereological criterion for being a simple. I call this view *the Brutal View*.[2]

My argument for the Brutal View is indirect. I argue that every reasonable answer to the Simple Question faces serious objections. Consequently, the Brutal View is the only view left standing. In Section 2, I motivate the quest to answer the Simple Question and briefly describe the space of possible answers. In Sections 3, 4, and 5, I present arguments against the competitors of the Brutal View. In Section 6, I respond to a possible worry about the Brutal View. In Section 7,

[1] The Simple Question was first raised by Ned Markosian in Markosian (1998a).

[2] The Brutal View is inspired by a related view that answers the Special Composition Question: the Brutal View of Composition. The Special Composition Question was first raised by Peter van Inwagen in van Inwagen (1990). (I will discuss the Special Composition Question momentarily.) On the Brutal View of Composition, see Markosian (1998b).

I present some final reflections on the Brutal View and the philosophical considerations that support it.[3]

2. THE SIMPLE QUESTION

Why care about the Simple Question? First, issues involving the nature of simplicity are not independent of other concerns in the metaphysics of material objects. Philosophical puzzles concerning material constitution have received a deserved share of the attention of contemporary philosophers; much of it focused on what Peter van Inwagen has dubbed the *Special Composition Question*, which is: What are the necessary and jointly sufficient conditions that some objects must meet in order to compose a single object? Ned Markosian, the philosopher to whom we owe gratitude for raising the Simple Question, notes the connection between these two questions in the following passage:

> simples are the basic building blocks that, when combined in various ways, make up all other objects. Thus it is natural to think that what we say about the nature of simples will have considerable bearing on what we say in response to the Special Composition Question. (Markosian 1998a: 214)

To see that Markosian is correct, let us consider a radical answer to the Special Composition Question called *Nihilism*.[4] Nihilism is the view that, necessarily, nothing is a composite object. Nihilism obviously conflicts with common sense concerning what objects exist, since one consequence of Nihilism is that there are no such things as tables, rocks, and living human organisms.[5] But the advocate of Nihilism at least agrees with common sense that there are material objects. However, given Nihilism, these must be mereological simples. If it turns out that nothing could satisfy what it takes to be a simple, or even if nothing in fact does satisfy what it takes to be a simple, then Nihilism is refuted.

[3] In what follows, I make few substantive assumptions about parthood in general: I assume that there is exactly one parthood relation defined on material objects, and that it is transitive and reflexive, but I do not assume any other controversial mereological thesis, such as the claim that no two objects can have exactly the same parts. I suspect that the arguments for the Brutal View would still work even if these meager assumptions were dropped.

[4] The name ''Nihilism'' was coined by Peter van Inwagen in van Inwagen (1990: 72–4).

[5] Strictly speaking, this is a consequence of Nihilism only given certain facts about the actual world. It is a fact about the actual world that, if there are tables, chairs, etc., then these objects have parts.

That the Special Composition Question is an interesting question is established by the fact that so many metaphysicians are interested in it. Since answering the Simple Question could shed light on the Special Composition Question, we should be interested in it as well.

Secondly, an answer to the Simple Question could help us decide whether *atomless gunk* is possible. An object is *gunk* just in case every part of it has proper parts. There are longstanding debates about whether gunk is possible. The question of the possibility of gunk is also relevant to answering the Special Composition Question; recently, Theodore Sider has argued that certain answers to the Special Composition Question are false because they rule out the possibility of gunk (Sider 1993). Additionally, Dean Zimmerman has argued that certain theories about the nature of masses are ruled out given the possibility of atomless gunk (Zimmerman 1995). If gunk is impossible, then these arguments have no force. An answer to the Simple Question may help us assess these arguments.

Another strange and putatively possible kind of object that has attracted its share of defenders is the *extended simple*. An object is an *extended simple* just in case it is extended in space (or spacetime) and yet lacks proper parts.[6] Speculation about the possibility of extended simples is not confined to philosophy. In a recent article, Mark Scala presents evidence that Isaac Newton believed that the fundamental objects of this world are extended simples (Scala 2002: 394).[7] And, more recently, in a popular book on string theory, the physicist Brian Greene seriously entertains the possibility that fundamental physics will imply the existence of extended simples:

> What are strings made of? There are two possible answers to this question. First, strings are truly fundamental—they are ''atoms'', uncuttable constituents, in the truest sense of the ancient Greeks. As the absolute smallest constituents of everything, they represent the end of the line... From this

[6] Friends of extended simples include Ned Markosian (Markosian 1998a), Neil McKinnon (McKinnon forthcoming), Josh Parsons (Parsons 2000), Mark Scala (Scala 2002), and Theodore Sider (Sider forthcoming).

[7] Here is the quote from Newton's *Opticks* that Scala discusses: ''It seems probable to me that God in the Beginning formed Matter in solid, massy, hard, impenetrable moveable Particles, of such Sizes and Figures, and with such other Properties; and in proportion to space, as most conduced to the end for which he formed them; and as these primitive Particles being Solids, are incomparably harder than any porous Bodies compounded out of them; even so very hard as never to wear or break in pieces; no ordinary Power being able to divide what God himself made in one first creation.''

perspective, even though strings have spatial extent, the question of their composition is without any content. Were strings to be made of something smaller, they would not be fundamental. (Greene 1999: 141)

If we had an answer to the Simple Question, this presumably would help us determine whether extended simples are possible.

So we should agree with Markosian that an examination of the Simple Question is relevant to an examination of the Special Composition Question. Moreover, it is an interesting question in its own right. Markosian should be commended for raising it.

Unfortunately, we are unable to answer the Simple Question satisfactorily. I will argue that there is no correct, finitely statable, non-trivial answer to the Simple Question. In short, the Brutal View of Simples is true.

It is not part of the Brutal View that there are no informative *necessary* conditions on being a simple. In fact, I believe that there are. But there are no informative conditions that are *both* necessary and sufficient for being a simple and finitely statable. I also distinguish the Brutal View from the claim that, for any simple *S*, it is a *brute fact* that *S* is a simple. We can call the latter view *the Brutal View of Facts about Simplicity* (BFS).[8] If there are informative sufficient conditions for being a simple, then BFS is false. But, as long as these sufficient conditions are not necessary conditions (or not finitely statable), then the Brutal View is unthreatened by the falsity of BFS.[9]

[8] The distinction between the Brutal View and BFS is perfectly similar to the distinction drawn by Markosian between the views he calls the Brutal View of Composition and the Brutality of Compositional Facts (Markosian 1998b).

[9] I thank Ben Caplan and Cody Gilmore for helpful discussion on this point. Interestingly, Markosian himself ignores the difference between the Brutal View and BFS when arguing against the Brutal View. Markosian writes:

For the record, here is what seems to me like a good argument against The Brutal View of Simples. A pointy object would have to be a simple. Moreover, such an object would be a simple in virtue of being pointy. (Either because The Pointy View of Simples—according to which x is a simple iff x is a pointy object—is true, or because MaxCon is true, or because some other ''topological'' theory of simples is true.) But the Brutal View of Simples implies that even if a pointy object were a simple, this would not be the case in virtue of its being pointy; instead, according to The Brutal View of Simples, the fact that some pointy object is a simple would have to be a brute fact about that object. Therefore, The Brutal View of Simples is false. (Markosian 2004: 333 n. 2)

There are two problems with Markosian's argument against the Brutal View. First, as already mentioned, he conflates the Brutal View with BFS. If he's hit something, he's hit the wrong target. Secondly, Markosian assumes that being point-sized is sufficient for being a simple. But, as I will argue in Section 3, this is not the case.

It is also not part of the Brutal View that there are no features that are contingently correlated with being a simple. It is my hope that there are. But, whatever they are, it is up to empirical science and not a priori philosophy to discover them.

Still, it is unfortunate that the Brutal View is true. For the Brutal View sheds no light on the question of whether atomless gunk is possible, whereas other answers to the Simple Question appear to. And the Brutal View provides us with no help with the question of whether extended simples are possible, whereas other answers to the Simple Question have something to say about the possibility of extended simples. The Brutal View does not tell us that atomless gunk or extended simples are possible, and it does not tell us that they are impossible.

Although the Brutal View of Simples is a dissatisfying answer to the Simple Question for these reasons, I believe that there is a compelling reason to embrace it: the competitors to the Brutal View face problems serious enough to warrant rejecting them. If this is the case, the Brutal View is the only game in town.

What are the competitors to the Brutal View? As I see things, the other main contenders can be divided into three classifications, each with two sub-headings. They are as follows:

(A) *Spatial Accounts*

- (1) *The Pointy View of Simples* (PV)
- (2) *The Maximally Continuous View of Simples* (MaxCon)

(B) *Fundamentality Accounts*

- (3) *The Instance of a Fundamental Property View of Simples* (Instance)
- (4) *The Independence View of Simples* (Independence)

(C) *Indivisibility Accounts*

- (5) *The Physically Indivisible View of Simples* (PIV)
- (6) *The Revised Metaphysically Indivisible View of Simples* (MIV)

In Section 3, I present and argue against the spatial accounts.[10] In Section 4, I present and argue against the fundamentality accounts. In Section 5, I present and argue against the indivisibility accounts. This will complete my case for the Brutal View of Simples.

[10] Markosian also presents several arguments against the Pointy View in Markosian (1998a: 216–19).

3. SPATIAL ACCOUNTS OF SIMPLICITY

As the name suggests, spatial accounts of simplicity appeal to spatial features to provide a criterion of simplicity. The two most promising are:

> **The Pointy View of Simples (PV)**: necessarily, *x* is a simple if and only if *x* is a point-sized object.
> **The Maximally Continuous View of Simples (MaxCon)**: necessarily, *x* is a simple if and only if *x* is a maximally continuous object.[11]

According to the Pointy View, simples are all and only point-sized objects. If you want to make a simple, create a point-sized object. (Don't ask me how to do that!) The Pointy View is probably the traditional view of the nature of simples.

The Pointy View has two interesting features. First, if material objects without point-sized parts are possible, then the Pointy View implies that gunk is possible (Markosian 1998a: 216). Secondly, the Pointy View clearly implies that extended simples are impossible.

MaxCon is not for traditionalists! Given MaxCon, if you want to make a material simple, here is the recipe you should follow. First, pick the region of space that you want the simple to exactly occupy. Let us call that region ''R''. If R is a continuous region of space, then proceed to the next step. Otherwise, start again. Assuming that R is a continuous

[11] MaxCon is the view that Markosian endorses in Markosian (1998a). Markosian employs the following definitions in the explication of his view:

1. Object O *occupies* region R $=_{df}$R is the set containing all and only those points that lie within O.
2. *O* is *spatially continuous* iff *O* occupies a continuous region of space.
3. R is *continuous* $=_{df}$ R is not discontinuous.
4. R is *discontinuous* $=_{df}$ R is the union of two non-null separated regions.
5. R and R′ are *separated* $=_{df}$ the intersection of either R or R′ with the closure of the other is null.
6. The *closure* of R $=_{df}$ the union of R with the set of all its boundary points.
7. p is a *boundary point* of R $=_{df}$ every open sphere about p has a non-null intersection with both R and the complement of R.
8. R is an *open sphere* about p $=_{df}$ the members of R are all and only those points that are less than some fixed distance from p.
9. The *complement* of R $=_{df}$ the set of points in space not in R.
10. *x* is a *maximally continuous object* $=_{df}$ *x* is a spatially continuous object and there is no continuous region of space, R, such that (i) the region occupied by *x* is a proper subset of R, and (ii) every point in R falls within some object or other.

Markosian borrows (2)–(9) from Cartwright (1975). (Richard Cartwright uses ''connected'' and ''disconnected'' instead of ''continuous'' and ''discontinuous''.)

region of space, completely fill R with matter; make sure that there is no subregion of R where matter cannot be found. Finally, make sure that R is not part of some larger continuous region of space that is also filled with matter. If it is not, then R now contains a material simple.

Presumably, R can be any size or any shape; the only constraint on R is that it be occupiable by a material object. Given MaxCon, there can be extended simples of any shape or size.[12] What about gunk? If MaxCon is true, then gunk is impossible.[13] MaxCon is a stunningly unorthodox answer to the Simple Question.

The main argument against spatial accounts of simplicity is based on the possibility of co-located point-sized objects.[14] Two objects are co-located if they exactly occupy the same region of space (at the same time).[15] The argument is as follows: (1) co-located point-sized objects are possible; (2) if co-located point-sized objects are possible, then mereologically complex point-sized objects are also possible. But then both the Pointy View and MaxCon are false.[16]

One way to motivate premise (1) is to appeal to the conceivability of co-location. We can form a clear and distinct conception of co-located material objects; they are conceivable. This gives us a reason to believe that they are possible.[17] For example, we can imagine two different kinds of stuff that are capable of interpenetrating. But we need not base

[12] More cautiously, MaxCon does not imply that there are restrictions on the shape or size of extended simples. There might be other restrictions on the shape or size of material objects that are consistent with MaxCon.

[13] Hudson argues this in Hudson (2001: 84–7).

[14] I thank Ryan Wasserman for pressing me on this point. See Wasserman (2003) where he also discusses this objection. Markosian acknowledges this worry, which he credits to Theodore Sider. See Markosian (1998a: 217 n. 20). Other arguments against MaxCon (but not PV) can be found in McDaniel (2003).

[15] Two points should be stressed. First, co-location in this sense must be distinguished from complete mereological overlap. Two objects completely overlap each other if and only if the two objects have exactly the same parts. Secondly, regions and material objects form distinct ontological categories. If we drop this assumption, the argument from co-location is undercut. There is no way to make sense of co-located regions of spacetime. I think that this is a reason to reject the reduction of material objects to spacetime regions, but others may differ. I thank Carl Matheson for helpful discussion of this point.

[16] Since MaxCon implies that any point-sized material object is a simple, the possibility of mereologically complex point-sized objects also refutes MaxCon.

[17] Perhaps this is a defeasible reason, if conceivability does not entail metaphysical possibility. But, nonetheless, it does provide us with (as of yet undefeated) evidence that they are possible. On the relation between conceivability and possibility, see the fine collection of papers published in Gendler (2002).

the case for the conceivability of co-located objects on the strange thought experiments of a philosopher. There is an interesting debate in the philosophy of quantum mechanics about whether *bosons*, a kind of fundamental particle, are counter-examples to the Identity of Indiscernibles.[18] Bosons are counter-examples to this law only if two or more of them can be at the same place at the same time. Peter Simons, in a recent paper on the bundle theory of objects, makes the point nicely:

> *Fermions*, which include electrons, are characterized by [properties] which obey the *Pauli Exclusion Principle*: no two fermions can be in exactly the same state. Thus the reason that a helium atom may have two electrons in its innermost shell is that their spins are in opposite directions, so they differ in one [property] (maybe a second-order [property]: spin-*direction*).... The other sort of particles are *bosons*. They do not obey the Pauli Principle, and so *two or more bosons can be in the same state at the same time, in particular they can be in the same place at once and not differ in any* [property] *at all*. If electrons were bosons, they could all three occupy the same space around a lithium nucleus. The most familiar bosons are photons, and it is their superposability in large numbers that makes lasers possible. (Simons 1994: 379–80)[19]

I am no expert on quantum physics, so I am unable to evaluate Simons's claim here. But I am not trying to argue that co-located objects are *actual*. What this example shows is that co-located material objects are not merely conceivable, but that a tremendously detailed conception of them has been formed: co-located objects play a role in the interpretation of certain physical theories. It might be that at the end of the day speculative physics will postulate co-located material objects. It seems to me that we should not disregard this possibility a priori. That both spatial accounts of simplicity do eliminate this possibility a priori is problematic.

Finally, there is the argument from systematic modal metaphysics: the mere metaphysical possibility of co-located objects follows from familiar Humean principles involving the denial of necessary connections

[18] For interesting discussions about this issue and the question of whether bosons violate the Identity of Indiscernibles, see Cortes (1976), Barnette (1978), Ginsberg (1981), and Teller (1983).

[19] I have emphasized the relevant part of the quote; also, in the original passage, Simons talks about tropes, whereas I have substituted the word "property" for "trope" uniformly. I don't think this change makes a difference in this context.

between distinct existences. Suppose two point-particles are approaching each other at a rapid clip. If co-located material objects are impossible, then they must swerve out of each other's way. Or they must stop dead in their tracks. Or one of them must spontaneously disintegrate. Some event must occur in each world that prohibits them from occupying the same space. There is a *de re* necessary repulsion between these two objects. The price of denying the possibility of co-located objects is accepting brute *de re* modal facts like these. The price is too high.[20]

The state of affairs in which an object *x* occupies a particular region of space *R* (at *t*) is distinct from the state of affairs in which an object *y* occupies the same region (at the same time). From the fact that the first state of affairs obtains, we can infer nothing about the location of *y*. Both states of affairs obtain contingently. If any recombination of distinct, contingent states of affairs yields a genuine possibility, as I am inclined to hold, then there are possible worlds at which both *x* and *y* occupy *R* (at *t*).[21]

Why believe premise (2), which says that if co-located objects are possible, then so are objects composed of them? Suppose that in some possible world two point-sized objects occupy the same region of space. Then there could be a thing made out of those objects. For example, suppose the two objects always move together because they are held by a fundamental physical force. Surely there are possible worlds in which the laws of nature guarantee this sort of interaction. If this scenario arose, we would say that the two objects were ''joined together'', ''bonded'', or ''fused''. In such a case, one would be hard pressed to say that they do not compose something. I suspect only the mereological nihilist could resist

[20] In response to this argument, David Robb suggested to me that a Humean could say that it is analytic that material objects do not interpenetrate; co-located objects are by definition not material objects. If co-located material objects are impossible for this reason, we still need to explain why they seem to be conceivable. Perhaps what we are envisioning is what David Robb called *phony matter*: non-material objects that have the same effects on our sensory states as do material objects. I think this suggestion won't work, for *material object* is an ontological category. If *x* belongs to an ontological category *C*, then *x* has the *intrinsic* property *being a C-entity*. So every duplicate of a material object is itself a material object. But our allegedly co-located material objects are duplicates of genuine material objects. So they themselves are material objects. So it is not analytic that material objects cannot interpenetrate.

[21] For defenses of these principles, see Armstrong (1989), Armstrong (1997: 148–84), and Lewis (1986).

this pressure. But, if the objects do compose something, then this composite object is a counter-example to the spatial accounts of simplicity.[22]

One more remark on co-location before I move on: I am inclined to accept that the following is a *sufficient* condition for being a simple: being point-sized and not co-located with any other point-sized object. Given this, some point-sized objects might be simples in virtue of having this property.[23] But it is not a necessary condition on being a simple. For the two co-located simples that composed the counter-example to the spatial accounts are still simple, despite their being co-located.

4. FUNDAMENTAL ACCOUNTS OF SIMPLICITY

Both of the views that I wish to discuss here tie simplicity to some ultimate feature of objects. They are:

> **The Instance of a Fundamental Property View of Simples (Instance)**: *x* is a simple if and only if *x instantiates* a perfectly natural property.[24]
>
> **The Independence View of Simples (Independence)**: *x* is a simple if and only if it is metaphysically possible that *x* is the only material object that exists.

I will examine Instance first. In order to state Instance properly, I need to invoke some metaphysical machinery. I assume the existence of *perfectly natural properties*.[25] It is these properties that ground *objective similarity*: if two things instantiate the same perfectly natural property, then they are objectively similar in that respect; *duplicates* are objects such that there is a one-to-one correspondence between their parts that preserves perfectly natural properties (and perfectly natural

[22] Note that, if composition is unrestricted, then we do not need the additional supposition that the co-located objects are bonded in order to ensure that they form a composite object.

[23] I thank Ben Caplan for stressing this point.

[24] For the most part, I ignore questions concerning the nature of the properties in what follows, such as whether the properties are repeatable *universals* or are themselves particulars, i.e. tropes. On the issue of tropes vs. universals, see Lewis (1997b) and Simons (1994).

[25] On naturalness in Lewis's sense, see Lewis (1997a), Lewis (1997b), Lewis (1986: 60–1), Schaffer (forthcoming), Sider (2001), and Sider (1995).

relations). Whether two things are objectively similar is metaphysically independent of our interests, desires, beliefs, or classificatory schemes.[26]

Once we have the concept of a natural property, we can define other useful concepts (Lewis 1986: 62–3). Intrinsic properties are properties that never differ between duplicates; if A and B are duplicates and A has intrinsic property *F*, then so does B. External relations do not supervene on the qualitative character of their relata; however, they do supervene on the qualitative character of the fusion of the relata. External relations should be contrasted with *extrinsic* relations, which do not even supervene on the qualitative character of the fusion of their relata. An example of an extrinsic relation is ownership. Ownership does not supervene simply on the qualitative character of the owner and the owned; instead, it supervenes on that character taken along with the various social facts that accompany it.

The perfectly natural properties (and relations) are those that are both required and jointly suffice to provide a complete description of the world (Lewis 1986: 60). The distribution of every other property supervenes on the distribution of the perfectly natural properties (and relations); the perfectly natural properties (and relations) are the minimal supervenience base of every world.[27]

Instance ties together the concepts of simplicity and naturalness. According to Instance, an object is a simple if and only if it *instantiates* a perfectly natural property.

Instance is a theory about the nature of material simples that is also in principle capable of answering the Fully General Simple Question, which is: Under what circumstance is an entity of *any* ontological category a simple? We might think that, for any category of entity we care to include in our ontology, it makes sense to divide the entities in that category into those that are simple and those that are complex. Accordingly, it would be nice to have a unified and fully general account of what it is to be a simple *simpliciter*. Theories that characterize simples in terms of spatial (or spatiotemporal) concepts cannot provide a unified account of the nature of *all* simples. This is because not every entity has spatial or spatiotemporal features. Similar remarks seem to apply to accounts that characterize simples in terms of indivisibility.

[26] More generally, since objective similarity also comes in degrees, the degree to which a given property is natural is independent of our beliefs, desires, or interests.

[27] The fundamental qualitative properties I discussed in Section 2 form a subset of the perfectly natural properties.

However, the concept of having a natural property is not a concept that necessarily applies only to material objects, for it is possible that there are natural properties that are had by non-physical objects. For example, certain psychological properties might be perfectly natural. (In fact, I hold that this is the case.) Cartesian spirits, which are paradigmatic examples of non-physical objects, could have these properties. Accordingly, Instance is in a better position to provide a unified account of simplicity than either Spatial Account.

One initial worry about Instance stems from the fact that many, if not all, of the fundamental properties at the actual world are determinables.[28] For example, consider *rest mass*. It is reasonable to think that this is a fundamental property. However, objects such as my body, this table, and the planet have rest mass. Should I conclude then that all of these things are simples? Clearly not.

Strictly speaking, it is the *determinates* of rest mass that are the best candidates for being perfectly natural. So perhaps this worry arises only if *some*, but not *all*, of the determinates of mass are perfectly natural. We could call these determinates *the fundamental quantities* of rest mass if we liked. If the fundamental quantities of mass are had only by physical simples, whereas the non-perfectly natural determinates of rest mass are had by complex material objects, then this particular version of the objection would be circumvented.

There are two problems with this maneuver. First, it is not certain whether there are fundamental quantities of rest mass in this sense. So this move is risky. Secondly, and more damaging, it seems that this sort of maneuver does not work in other cases. Consider, for example, *charge*. *Being –1 charged* is a fundamental quantity of charge if any is. An electron, which is arguably a simple, has a charge of –1. However, consider a negatively charged isotope that has a charge of –1 because it has an extra electron. It has a fundamental quantity of charge and hence instantiates a perfectly natural property. This isotope is clearly not a simple.[29]

Perhaps a way around this problem is to claim that the isotope has a charge of –1 *derivatively*, that is, in virtue of the charge of its *parts*, and

[28] Jonathan Schaffer brought this argument to my attention.

[29] A third worry is that this move seems to violate an intuitive principle governing determinates and determinables: *p* and *q* are determinates of the same determinable *only if* *p* and *q* are equally natural properties.

so on for the other quantities. Similarly, one might hold that in some sense, I *inherit* the mass that I have from the mass of my parts; my mass is supervenient upon the mass of these objects, and likewise for the charge of the isotope. We could revise Instance so that it takes account of this intuition:

> **Instance***: *x* is a simple if and only if *x* instantiates a perfectly natural property non-derivatively.

However, to say that an object has a property derivatively is to say that it has the property in virtue of its parts having that property. So Instance* is actually circular; it violates one of the constraints on being an answer to the Simple Question, and hence (even if true) is no competitor to the Brutal View. (And, even if there is a way around this worry, Instance* would still face the previously discussed problems.)

Here is probably the best response to this worry: instead of moving to Instance*, the friend of Instance should instead distinguish the property of *having a net charge of –1* from the property of having a charge of –1.[30] An object has a net charge of –1 just in case the sum of the quantities of charge of its proper parts is equal to –1. According to this strategy, the isotope has a net charge of –1 but it does not have the property of having a charge of –1. If this strategy is viable, this kind of counter-example to Instance fails. It's not clear to me, however, that the composite object does not have the property of having a charge of –1 as well as the property of having a net charge of –1. So I am unsure whether this move is successful. On the assumption that this is an acceptable response to this kind of worry, I shall press on.

One might try to motivate Instance by appealing to the idea that there are many ways in which nature can be divided in hierarchies or levels.[31] One kind of hierarchy is *mereological*: some objects are parts of others, which in turn have parts of their own. This hierarchy terminates at the lowest level with mereological simples. Another kind of hierarchy is *qualitative*: some properties supervene on other properties, which in turn supervene on more basic or natural properties. This hierarchy terminates at the level of the perfectly natural properties. Or perhaps each of these hierarchies descends ''forever''; there are no mereological

[30] I owe this suggestion to Phillip Bricker.
[31] See Jonathan Schaffer (2003b) for a lengthy discussion of this picture.

simples or perfectly natural properties. But it is reasonable to hope that these hierarchies march in step, so that the simplest material objects enjoy the most natural properties. This is one motivation for Instance.

A second motivation for Instance is the intuition had by many that the properties of wholes are strongly dependent on the properties and relations of their proper parts. For some properties of wholes, that there is this kind of dependence is obvious, for example, the shape of a complex material object is fixed by the shapes of its parts and the spatial relations obtaining between those parts. But some people have the intuition that this kind of dependence holds for every intrinsic property of a whole. Perhaps they hold this view because they think they hold some form of the view that *composition is identity*.[32] If a whole just is its parts, then the properties of the whole should in some sense also be nothing more than the properties and relations of its parts. Given composition as identity, the intrinsic character of a whole is fixed by the characters and relations of its parts. The following principle is a way of formally stating this kind of dependence:

> (PWD): For every object *x* and all objects *y*s such that *x* is the fusion of the *y*s, and for all worlds *w1* and *w2*, if each of the *y*s has the same intrinsic properties in *w1* as it has in *w2*, and the *y*s stand in the same relations to each other in *w1* as they do in *w2*, then *x* has intrinsic property *F* in *w1* if and only if *x* has *F* in *w2*.[33]

In other words, given PWD, a whole cannot enjoy intrinsic variation across possible worlds unless either one of its proper parts enjoys intrinsic variation across possible worlds or its proper parts change with respect to the relations that they bear to each other.

I will now argue that, if you like PWD, you have some reason to like Instance. My first premise is a Humean principle to the effect that there are no necessary connections between the instantiations of the perfectly natural properties of contingent beings.[34] I formulate the Humean Principle as follows:

> (HP): If (1) *x* has *F* and *y* has *G*, (2) *x* and *y* are contingently existing material objects, and (3) *F* and *G* are perfectly natural properties, then

[32] On some of the ways of formulating the idea that composition is identity, see Sider (forthcoming).

[33] When I speak of relations here, I mean *external* relations.

[34] See Armstrong (1989) and Armstrong (1997).

there is a possible world in which both *x* and *y* exist, but in which *x* has *F* and *y* does not have *G*.

Informally, the instantiation of any perfectly natural property or relation is metaphysically independent of the instantiation of any perfectly natural property or relation.

Suppose that there is a complex material object *x* that instantiates a perfectly natural property *F*. Because *F* is perfectly natural, its instantiation is independent of the instantiation of other perfectly natural properties (for the Humean reason just given). So there is a possible world in which all of *x*'s proper parts have the same perfectly natural properties and stand in the same perfectly natural relations, but in which *x* does not instantiate *F*. Since *F* is perfectly natural, *F* is also an intrinsic property. Since *x*'s proper parts all have the same perfectly natural properties and stand in the same perfectly natural relations to each other, all of *x*'s parts have the same intrinsic properties. So there is a possible world in which all of *x*'s proper parts have the same intrinsic properties and stand in the same relations as they do in the actual world, but in which *x* differs intrinsically. So our assumption that a complex object has a perfectly natural property has led us to the conclusion that PWD is false.

So, at the very least, there is an argument from PWD for the claim that instantiating a perfectly natural property is *sufficient* for being a simple. One can also produce an argument for the claim that having a perfectly natural property is *necessary* for being a simple. Its premises are straightforward. First, every object, whether simple or complex, must have some intrinsic properties. Suppose that a simple *x* has an intrinsic property *P*. Either *P* is a perfectly natural property, or it supervenes on the perfectly natural properties and relations had by objects that are not identical with *x*, or *P* supervenes on the perfectly natural properties had by *x*. If the first disjunct is true, then *x* has a perfectly natural property. The second disjunct cannot be true, for if it were, then *P* would not be an intrinsic property; *P* would be an extrinsic property.[35] This leaves the third disjunct. Obviously, if the third disjunct is true, then *x* has a perfectly natural property. So, since every material object must have some intrinsic properties, then having a perfectly natural property is necessary for being a simple. If we conjoin

[35] Since *x* is a simple, any object to which *x* bears a relation is not a proper part of *x*.

these two results, we arrive at Instance: an object is a simple if and only if it instantiates a perfectly natural property.

At this point, one might naturally worry that the Humean Principle is too strong. The Humean Principle rules out cases in which the perfectly natural properties of a whole might be necessarily connected to the perfectly natural properties of some of its parts. One might then also worry that the principle is not supported by the familiar Humean intuition that there are no necessary connections between *distinct* existences. These intuitions could be seen to support only a weaker principle. Accordingly, one might want to relax HP and replace it with a weaker principle:

> (WHP): If (1) *x* has *F* and *y* has *G*, (2) *x* and *y* do not overlap, (3) *x* and *y* are contingently existing material objects, and (4) *F* and *G* are perfectly natural properties, then there is a possible world in which both *x* and *y* exist, but in which *x* has *F* and *y* does not have *G*.

These worries are misguided. The Humean Principle is not too strong, but WHP is too weak. To see this, consider a different sort of case that the Humean Principle correctly rules out but that WHP allows. Consider two properties, *F* and *G*, such that *F* and *G* are both perfectly natural, can only be instantiated by simples, and are necessarily connected in the sense that, necessarily, something has *F* only if something has *G*. (Something, however, can enjoy *G* without having *F*.) Clearly, this necessary connection is mysterious, and the sort of connection that any decent Humean will want to rule out. Since both *F* and *G* are perfectly natural, why can't something be *F* without being *G*? The mystery is not diminished when we learn that the things that have *F* always overlap (and are in fact identical with) some of the things that have *G*.

The Humean thinks that every perfectly natural property is a *distinct* existence, even if the things that instantiate them are not always distinct. And so there should be no necessary connections between perfectly natural properties, even if there are necessary connections between overlapping objects that instantiate them.

So an interesting case can be made for Instance given HP and PWD. I am inclined to endorse HP, but there is a good reason to reject PWD. This reason for rejecting PWD is also a reason to reject Instance.

I think that it is possible for mereologically complex objects to instantiate perfectly natural properties. I think this because I think

that some mereologically complex objects actually instantiate perfectly natural properties. Specifically, I am a mereologically complex material object who instantiates perfectly natural properties. I hold that certain mental properties, such as having a blue sensation or being in pain, are perfectly natural properties, or, at the very least, supervene on perfectly natural properties had by complex objects.

The argument that some phenomenal properties are perfectly natural is reasonably straightforward, but, of course, very controversial. The first premise is that there is a *zombie world*. A zombie world is a possible world that satisfies the following conditions: (1) every fundamental particle that exists in the actual world exists in the zombie world, (2) no fundamental particle exists in the zombie world that does not exist in the actual world, (3) every fundamental particle has the same intrinsic properties in the actual world as it has in the zombie world, (4) the fundamental particles stand in the same external relations to each other in the zombie world as they do in the actual world, and (5) nothing experiences episodes of phenomenal consciousness, such as having a blue sensation or feeling pain, in the zombie world.[36] I accept the first premise, because I seem able to conceive of a situation in which everything is just alike at the microscopic level, but in which no one enjoys qualitative experiences. (I also note that I am presupposing that every fundamental particle is a mereological simple and that there are no non-physical mereological simples such as Cartesian spirits.) The second premise is that phenomenal properties are *intrinsic* properties. I do not know how to argue for this claim; it seems intuitive to me, although I acknowledge that there is some controversy about whether it is true.[37]

These two premises imply that PWD is false. If they are true, the case for Instance has been undercut. Moreover, when supplemented with a third premise, they provide a reason to reject Instance. The third premise is this: if zombie worlds are possible, and phenomenal properties are intrinsic properties, then phenomenal properties are perfectly natural properties or supervene on perfectly natural properties had by

[36] On the possibility of zombies, see Chalmers (1996: 94–9). I note that the defender of the irreducibility of phenomenal properties to physical properties need not reject Instance, if she is willing to embrace a form of *panpsychism*, according to which phenomenal properties supervene on *proto-psychical* properties. For more on this interesting issue, see Chalmers (1996: 26–127).

[37] On this issue, see Merricks (2003) and Sider (2003).

mereologically complex objects. Since phenomenal properties are had by complex wholes, these three premises imply the falsity of Instance.

Why believe the third premise? Recall that the distribution of every qualitative property supervenes on the distribution of the perfectly natural properties and relations. So there can't be two worlds that differ qualitatively without differing with respect to some perfectly natural property or relation. A zombie world is a world that differs qualitatively from our world. So it must differ with respect to some perfectly natural property or relation. But it does not differ with respect to any of the perfectly natural properties or relations that are instantiated by the fundamental particles. So it must differ with respect to the perfectly natural properties had by some composite object.[38] So some composite object in the actual world must have a perfectly natural property that is not had by a composite object in the zombie world. So a composite object in the actual world has a perfectly natural property. This state of affairs is a counter-example to Instance.

I have presupposed that mereologically complex material objects are the bearers of phenomenal properties, and I stand by this presupposition. But one could maintain Instance if one rejected this claim. The existence of perfectly natural properties is not a problem for Instance if there exist mereological simples that instantiate them. In fact, one could argue for the existence of simple immaterial substances from the premises that (1) Instance is true, (2) being in pain is a perfectly natural property, (3) something is in pain, and (4) no material simple instantiates being in pain.

In general, Instance rules out the possibility of genuinely emergent properties. This is a reason to be concerned. Independently of concerns stemming from the philosophy of mind, it seems to me that we can conceive of situations in which perfectly natural properties are instantiated by mereologically complex objects. Suppose, for example, that physicists discover that bodies that appear to be particle-per-particle duplicates nevertheless behave differently when in the presence of a

[38] Strictly speaking, there is another alternative: the worlds may differ with respect to some perfectly natural relation instantiated by composite objects, upon which the phenomenal properties supervene. These composite objects would have to be parts of the objects that have the phenomenal properties on pain of these properties being extrinsic. On this alternative, zombie worlds do not provide a counter-example to Instance. However, I suspect that anyone who takes the possibility of zombies seriously will not be tempted by this alternative.

third kind of thing. That is, although A and B have the same subatomic structure, when in the presence of a third object clearly qualitatively different from A and B, effect E1 is produced when A is present, whereas effect E2 is produced when B is present. Suppose that these physicists observe a large number of instances of this kind of event. They accordingly divide bodies into sorts: those that behave like A, and those that behave like B. From a microphysical perspective, all of these bodies appear to be duplicates. It is reasonable to think that something else accounts for the difference in their behavior.[39] So, since nothing yet discovered at the level of microphysics does, that difference must be a difference at the macrophysical level. Some bodies must have a feature that others lack. In this kind of case, scientists would be justified in postulating natural properties that are had by macrophysical wholes, not their parts. Instance, however, rules out the possibility of macrophysical wholes enjoying fundamental properties a priori.

And we need not rest the case for the conceivability of emergent properties on the bare bones thought experiment just given. For on some interpretations of quantum mechanics, the quantum state of the universe is a perfectly natural property assigned to a complex whole: the entire physical universe. This has recently been argued by Jonathan Schaffer (Schaffer, forthcoming). In a passage by Tim Maudlin (quoted also by Schaffer), Maudlin notes:

> In quantum theory, then, the physical state of a complex whole cannot always be reduced to those of its parts, or to those of its parts together with their spatiotemporal relations, even when the parts inhabit distinct regions of space. (Maudlin 1998: 56)

As Schaffer puts it, ''In other words, *mereological supervenience fails*. The properties of entangled wholes do not supervene on the intrinsic properties and arrangements of their parts'' (Schaffer, forthcoming).

Again, I'm not a quantum physicist, so I'm in no position to evaluate whether the universe or any complex part of the universe enjoys this kind of quantum entanglement. But I'm not trying to argue (in this

[39] Although perhaps we are not *required* to think that this is the explanation. An alternative explanation is that the laws of nature at this world are indeterministic. (I thank C. L. Hardin for bringing this point to my attention.) Of course, we are not required to think this either. My point is that there are possible situations in which we could be justified in positing genuinely emergent properties.

context) that emergent properties are *actual*. I'm arguing that they are conceivable. Their conceivability is demonstrated by the fact that actual respectable scientific theories appeal to emergent properties. This provides a reason to think that emergent properties are metaphysically possible, which is sufficient to eliminate Instance as an answer to the Simple Question.

This completes my case against Instance. I will now discuss:

> **The Independence View of Simples (Independence)**: *x* is a simple if and only if it is metaphysically possible that *x* is the only material object that exists.

The basic motivation for Independence is this: since simples are the fundamental building blocks of reality, they can be fully recombined. The idea that simples can be fully recombined finds its clearest statement in the work of D. M. Armstrong in *A Combinatorial Theory of Possibility*. In that book, Armstrong develops a Humean account of modality that implies that any simple can coexist with any other simple and that if something is a simple, then it is metaphysically possible for it to exist alone (Armstrong 1989: 37–48, 61–2).

I endorse the Humean program in modal metaphysics, so I will not challenge the claim that, if something is a simple, then it is metaphysically possible that it is the only material object that exists. However, this is not to say that the claim will be acceptable to all. Many philosophers claim that objects have their *origins* essentially. Suppose that an electron was created as a result of the big bang. Suppose that the big bang would not have happened had there not been an initial singularity, that is, a point-sized object of enormous density. If objects have their origins essentially, then our electron could not have existed unless that singularity had also existed. But nevertheless the electron is still an excellent candidate for being a simple. So Independence will be accepted only by those who reject certain kinds of essentialist views.

My first worry about Independence is that it seems that some composite objects could satisfy the right hand-side of the biconditional. For consider a composite object that could have been a simple.[40] If this

[40] Markosian discusses the inverse of this, specifically, the possibility that a simple become a composite in Markosian (1998a: 221). Admittedly, it is controversial whether these alleged possibilities are genuine. For example, mereological essentialists will deny that these possibilities are genuine.

object could have been a simple, then it, like other simples, could have been the only material object in existence.[41] But then it satisfies the right-hand side of the biconditional. But, since it is not actually a simple, Independence is false.

The advocate of Independence can avoid this worry by revising her view as follows:

> **Independence***: x is a simple if and only if there is a possible world w at which (1) x is the only existing material object and (2) x instantiates an intrinsic property P at the actual world if and only if x instantiates P at w.

Independence* avoids the counter-example that plagued its ancestor. Perhaps a composite object could have been a simple. But any object has a different intrinsic character in worlds in which it is a simple than in worlds in which it is complex.

I am inclined to think that Independence* is true. I think that Independence* provides necessary and sufficient conditions for being a simple. My worry is that Independence* violates the non-circularity requirement on being an answer to the Simple Question. Independence* appeals to the notion of an intrinsic property, and this concept is partly mereological. Recall the definition of ''intrinsic property'': a property is intrinsic if and only if it never differs between duplicates. Now recall that the analysis of duplication also appealed to the concept of parthood: x and y are duplicates if and only if there is a one-to-one correspondence between their *parts* that preserves perfectly natural properties and relations. So Independence* may provide necessary and sufficient conditions for being a simple, but Independence* is consistent with the Brutal View.[42]

[41] I assume here a modal logic at least as strong as S4.

[42] It is worthwhile to see a second attempt to salvage Independence. Consider *Independence***, according to which an object x is a simple if and only if there is a possible world w at which (1) x is the only existing material object and (2) x instantiates a perfectly natural P at the actual world if and only if x instantiates P at w. (This version was suggested to me by Ben Caplan.)

Since the concept of a perfectly natural property is not a mereological concept, Independence** is not circular. However, I think we can construct a possible counter-example to Independence**. Consider a possible world w in which a composite object o does not instantiate any perfectly natural properties. Suppose that o could have been a simple such that for any property p it instantiates, there is a property q such that q is more natural than p. In such a world, o does not instantiate a perfectly natural property either. If such a case is possible, then Independence** implies that o is actually a simple, which is false.

A deeper worry along the same lines is that it may turn out that *modal* concepts are partly mereological as well. If this is the case, Independence* is doubly guilty of violating the non-circularity requirement on being an answer to the Simple Question. Consider, for example, David Lewis's account of modality (Lewis 1986). According to Lewis, a proposition is possibly true just in case it is true at some possible world. A possible world is the *mereological sum* of maximally spatiotemporally related entities. So, on Lewis's view, mereological terms appear in the analysis of modality. If a view like Lewis's is correct, then any putative answer to the Simple Question that appeals to modal concepts violates the circularity requirement and hence is not a competitor to the Brutal View.

Obviously, I can't prove here that the correct analysis of modality must appeal to mereological concepts. The chapter is long enough as it is! Suffice it to say that this worry is one that a friend of Independence should take seriously.

This completes my case against Independence.

5. INDIVISIBILITY ACCOUNTS OF SIMPLICITY

As the name suggests, Indivisibility Accounts appeal to the concept of *indivisibility* when answering the Simple Question. The main idea behind the Indivisibility Accounts seems to be this: simples are things without proper parts, and hence cannot (in some sense) be split apart. So things that can be divided must not be simple. On the other hand, things that cannot be divided must have no parts to be separated, and so must be simple.

Markosian distinguishes two versions of the Indivisibility Account, which he calls:

> **The Physically Indivisible View of Simples (PIV)**: *x* is a simple if and only if it is not physically possible to divide *x*.
>
> **The (Revised) Metaphysically Indivisible View of Simples (MIV)**: *x* is a simple if and only if it is not metaphysically possible to divide *x* without first changing *x*'s intrinsic properties. (Markosian 1998a: 220–1)[43]

[43] Markosian also discusses an unrevised version of MIV; since I believe the argument he makes against it is sound, I will not discuss it here.

My first worry about the Indivisibility accounts is that they appear to violate the non-circularity condition on being an answer to the Simple Question, and hence, even if true, will not be competitors to the Brutal View. It seems that the concept of divisibility cannot be explicated without appealing to mereological concepts in the explication. Consider the following analysis of *divisibility*:

> (D1): *x* is *divisible* if and only if it is possible that there are objects *y* and *z* such that (1) *x* is composed of *y* and *z* and (2) the union of the regions occupied by *y* and *z* is discontinuous.

D1 has two interesting features. First, it does not imply that divisible objects have proper parts, but it does imply that divisible objects possibly have proper parts. Secondly, D1 implies that divisible objects can survive division. A different account of divisibility that does not have these features is:

> (D2): *x* is divisible if and only if there are objects *y* and *z* such that (1) *x* is composed of *y* and *z* and (2) it is possible that the union of the regions occupied by *y* and *z* is discontinuous.

D2 implies that divisible objects have proper parts, but it does not imply that divisible objects can survive division.

Notice that both accounts of divisibility employ mereological concepts. So any account of simplicity that employs the concept of divisibility and then explicates this concept along the lines of D1 or D2 violates one of the conditions on being an answer to the Simple Question by appealing to mereological concepts in the right-hand side of the answer. Without a non-circular account of *divisibility*, the divisibility accounts are not competitors to the Brutal View.

Perhaps we could appeal to the concept of *matter* or *stuff* when giving an account of divisibility. (Markosian argues that the MaxConist needs to appeal to the persistence of *matter* in order to account for the qualitative heterogeneity of extended simples in Markosian (1998a: 223–6). Perhaps the advocate of the Indivisibility Accounts should also appeal to *stuff* to solve some of her theoretical problems.) Consider the following account of divisibility:

> (DM): *x* is divisible if and only if there is some matter *M* such that *M* ''makes up'' *x* and it is possible that *M* occupies a discontinuous region.

In order for this account to be intelligible, we need a clear account of the notion of matter and what the nature of the relation between an object and the matter that makes up or constitutes the object. So a brief digression is necessary.

Some philosophers hold that the material world is fundamentally a world of *stuff*, not of *things*. One friend of stuff is Michael Jubien, who writes:

the world does not come *naturally* divided into a definite array of discrete things. Instead, it consists of ''stuff'' spread more or less unevenly and more or less densely around space-time. . . . I am taking it as a fundamental ontological doctrine that the raw material of the physical universe is *stuff*, not *things*, and that the organization of (some of this) stuff into things is done by *us*. (Jubien 1993: 1–2)

Jubien claims that a complete description of the physical universe need not employ the concept of a thing (Jubien 1993: 2). Andrew Cortens says something similar in his sympathetic description of the stuff ontology:

According to [the stuff ontology], reality is to be thought of, not as a collection of objects, but rather, as being made up of stuff of various kinds. . . . On this view, mass terms serve as the best vehicle for representing reality in a perspicuous way. [Stuff ontologists] will resist any attempt to recast ''stuff-talk'' into standard object idioms. Any attempt to do so, however ''elegant'' from a purely formal point of view, they will view as being a move away from, rather than toward, greater perspicuity. In view of this, it seems reasonable to say that the stuff-ontologist endorses a picture of reality which excludes objects. (Cortens 1997: 46–7)

I take it that the central doctrine of the stuff ontology is that truths about the properties and relations of things—if there are any such truths—always supervene on the truths about the properties and relations had by various stuffs. If we wish to assert these truths in a maximally perspicuous way, we should use sentences employing mass terms, not count nouns.[44]

[44] In a similar vein, Theodore Sider writes:

It is important to be clear on how radical this view must be, if it is to be a genuine alternative to a thing-ontology. Some philosophers talk as if they defend a stuff-ontology, when they really just believe in things in stuff's clothing: 'The world consists of quantities of stuff; we can decide to interpret thing-quantifiers as ranging over any of the quantities of stuff we choose. One could use thing-quantifiers to range only over small bits of stuff, in which case the nihilist is right. Or one could use the thing-quantifiers to range over *all* the

I reject the stuff ontologist's attempts to eliminate things or reduce talk about things to talk about stuff. Now a friend of stuff need not endorse wholeheartedly the stuff ontology; she need only hold that some truths about stuffs are not reducible to any class of truths about things. Some facts about stuff are basic facts about the world, even if not all of the basic facts about the world are about stuff. This I reject as well: the truths about the properties and relations stuffs bear—if there are any such truths—supervene on more fundamental truths about the properties and relations had by things. The world is a world of things, not stuff.[45] (And, moreover, every world is a world of things, not stuff.)[46]

Suppose that the properties and relations instantiated by mereologically complex objects supervene on the properties and relations instantiated by mereological simples. That is, suppose that, once we have fixed the properties and relations of all of the material simples, we have fixed the properties of and relations of every complex object. If this is the case, and the world is a world of things, then there is nothing else that the properties and relations of the simples supervene on. There is no ''fundamental stuff'' that (i) ''constitutes'' or ''makes up'' these simples, and (ii) is such that the properties and relations of the simple objects supervene on the properties and relations of this stuff. This is one metaphysical consequence of the doctrine that the world is a world of things.

quantities of stuff, in which case there exists scattered objects.' In fact, this view assumes that the world is a world of things: *quantities* of stuff.... A genuine no-conflict stuff ontologist must claim that a truly fundamental description of the world must completely eschew a thing-language. This requires completely eschewing the usual quantifiers and variables—the backbone of contemporary logic.... A whole new language must be developed. Somehow, 'quantifiers' over stuff must be introduced without slipping into talk of things; somehow language must be invented to express all the facts about the world we take there to be, while not slipping into thing-language in disguise. (Sider 2001: pp. xvii–xviii)

[45] For a contrary view, see Markosian (forthcoming), in which he defends a mixed ontology including both things and stuff.

[46] This doesn't mean that we must eschew mass terms. That would be an entirely inappropriate response to the claim that the world is a world of things. We are still allowed to say, ''Some water is wet'' and ''More mashed potatoes is always better than less''. But the truth-values of these sentences are determined by facts about things. Specifically, that some portions of water—which are *things*—are wet suffices to ensure the truth of ''Some water is wet''; likewise, the fact that it is always better to receive a larger portion of mashed potatoes than a smaller portion entails the proposition expressed by ''More mashed potatoes is always better than less''.

I've argued elsewhere against including stuff into our ontology along with things; I won't repeat the argument here.[47] Suffice it to say that this way of formulating the notion of divisibility carries with it a high ontological price, which some may not be willing to pay.[48] There are other worries about the divisibility accounts, to which I now turn.

My second worry is that the physical divisibility account seems to be a non-starter. My worry stems from the fact that *being physically indivisible* seems to be an *extrinsic* property. An object might be physically indivisible in world *w* and yet be physically divisible in a world with different natural laws. Yet that object may have the same intrinsic nature in both worlds.

But *being a simple* is not an extrinsic property. It is provably an intrinsic property. Suppose that *x* and *y* are duplicates and that *x* is a simple. Since *x* and *y* are duplicates, there is a one-to-one correspondence between their parts that preserves perfectly natural properties. But then there is a one-to-one correspondence between their parts. So *y* is a simple. So simplicity is preserved by duplication. So being a simple is an intrinsic property.

If two properties are necessarily co-extensive, then one of them is an intrinsic property if and only if the other property is. *Proof*: assume *P* and *Q* are necessarily co-extensive. Then *P* never differs between duplicates if and only if *Q* never differs between duplicates. Intrinsic properties are properties that never differ between duplicates. So *P* is intrinsic if and only if *Q* is intrinsic.

Since *being physically indivisible* is an *extrinsic property* and *being a simple* is an intrinsic property, and since it is impossible for an extrinsic property to be necessarily co-extensive with an intrinsic property, PIV is false.

I will now discuss the Revised Metaphysically Indivisible View of Simples. Markosian writes this about MIV:

Unfortunately, [MIV] is equivalent to the Pointy View of Simples. For it seems clear that all and only pointy objects would satisfy the right-hand side of the

[47] See McDaniel (2003).

[48] Markosian argues that any friend of extended simples must accept both stuffs and things in her ontology (Markosian, forthcoming). Although I am dubious about this claim (see Gilmore (forthcoming) for a convincing argument that postulating stuff does not eliminate worries about extended simples), it is interesting to note that the friends of the Indivisibility Accounts of simples may also need to make this commitment, despite the apparent fact that Indivisibility accounts are hostile to the possibility of extended simples.

bi-conditional.... Thus the above objections to the Pointy View of Simples would also apply equally well against this view. (Markosian 1998a: 221)

I think that this is mistaken, although it is hard to tell, since we do not have a clear account of the notion of metaphysical divisibility. However, if we operate with our shaky but intuitive grasp of this concept, I think we can agree that some possible point-sized objects are metaphysically divisible. Consider a point-sized object that is composed of two other point-sized objects. (I argued that this kind of case is possible in Section 2.) This object seems to be divisible, for it is possible for its parts to be in distinct regions of space. The Pointy View incorrectly implies that this object is a simple; MIV does not have this implication. MIV and the Pointy View are not equivalent.[49]

So Markosian's reason for rejecting MIV doesn't work. Nonetheless, I think there is a good reason to reject MIV as an answer to the Simple Question: MIV faces the circularity worry in a second (and perhaps third) guise. In addition to appealing to the concept of divisibility, MIV also appeals to the concept of an intrinsic property. And this concept is partly mereological, as I argued in the previous section. So MIV is guilty twice-over of sneaking mereological concepts into the analysis of simplicity. (It may be thrice guilty if modal concepts are also partly mereological, as I suggested they might be in the previous section.)

This completes my case against MIV.

6. UNKNOWABLE SIMPLES?

Here is a commonly told story: we used to think that chemical atoms were also atoms in the original sense, that is, mereological simples. But then we discovered that atoms are not mereological simples: we discovered that atoms are composed of a nucleus and the electrons in the outer-shells surrounding the atom. Perhaps there is further structure yet to be discovered? As Jonathan Schaffer writes:

> Indeed, the history of science is a history of finding ever-deeper structure. We have gone from ''the elements'' to ''the atoms'' (etymology is revealing), to the subatomic electrons, protons, and neutrons, to the zoo of ''elementary particles,'' to thinking that hadrons are built out of quarks, and now we are

[49] This example also shows that it is not the case that something is divisible if and only if it is extended in space.

sometimes promised that these entities are really strings, while some hypothesize that the quarks are built out of preons (in order to explain why quarks come in families). Should one not expect the future to be like the past? (Schaffer 2003b: 503)

There are two related worries that this picture seems to generate for the Brutal View. First, it seems that we often discover that certain objects are not simple. We might worry that, if the Brutal View of Simples is true, then we could not discover whether these objects were simple. What criterion could we use to rule that some object is not a simple if the Brutal View of Simples is true?

Secondly, it seems as if the search for the fundamental physical objects, by which I mean material mereological atoms, is one of the large projects in the history of physics. But it is hard to see how we could hope to succeed in this endeavor—*even if the world does divide without remainder into mereological* atoms—if the Brutal View of Simples is true.

Many answers to the Simple Question do not face these worries. For example, consider MaxCon. We have discovered that the nucleus of, for example, a hydrogen atom is actually some distance apart from the electron. This means that a hydrogen atom is not a maximally continuous object. So MaxCon correctly implies that a hydrogen atom is not a simple. Moreover, MaxCon can guide us in our search for the fundamental level: if we wish to find out which objects are mereological atoms, we should find out which objects are maximally continuous.

Similarly, the Pointy View can guide us in our search for the fundamental level: if we wish to find out which objects are mereological atoms, we should find out which objects are point-sized. Many of the other answers to the Simple Question seem to have this feature as well. Instance, for example, implies that the fundamental physical objects are also the basic bearers of perfectly natural properties. So once we discover those properties on which all others supervene, we will have discovered the true atoms of the world as well.

But it seems that the Brutal View of Simples cannot provide any guidance in our search. How then, given the Brutal View of Simples, could we ever know that our search had come to a conclusion? Perhaps some objects really are the true elements of the world. But the Brutal View of Simples won't tell us that they are. How then could we know that they are? We might be tempted to say that discovering what

objects are the actual mereological simples is the job of scientists, not philosophers, but without some idea of what they are looking for, how will they know when they have found it?

It is true that the Brutal View of Simples does not provide this sort of guidance. But this does not mean that the Brutal View of Simples is inconsistent with other principles that could provide us with aid in our quest to discover the true atoms of the world. As I noted in Section 2, the Brutal View of Simples is consistent with the existence of necessary conditions on being a simple, provided that these conditions are not both sufficient and informative. Similarly, it is consistent with the Brutal View of Simples that there are sufficient conditions for being a simple; as long as these sufficient conditions are not both also necessary and informative the advocate of the Brutal View of Simples need not do without them.

7. CLOSING REMARKS

The Brutal View is a somewhat unsatisfying answer to the Simple Question. But, if the most plausible alternatives to the Brutal View fail, a reasonable hypothesis why they fail is that simples per se have no nature.

Many of the premises employed in the arguments against the Brutal View's rivals are controversial. And, since the direct argument for the Brutal View is an argument via elimination, the case for the Brutal View is somewhat shaky. Specifically, I am aware that some of the modal principles I cherish and employ throughout this chapter—specifically, recombination principles conceived in a broadly Humean spirit—are not cherished by all.[50]

In the closing remarks of the paper in which he first raised the Simple Question, Markosian makes this observation:

> Many of the above reasons in support of MaxCon, as well as the arguments I have given against MaxCon's rivals, are based on appeals to intuitions about what should be said concerning various possible cases. Such ''modal intuitions'' are notoriously difficult to defend. I understand that many philosophers who read this paper will not be convinced by my arguments, precisely because they do not share my modal intuitions about the relevant

[50] But they are cherished by enough philosophers to ensure that arguments employing them will not be without interest.

cases. But this is a common phenomenon, especially in discussions of fundamental metaphysical issues, and it would be a mistake to expect anything else. I hope that the arguments of the paper will nevertheless be valuable even to those who do not share my modal intuitions. For it can be worthwhile to see what there is to be said for a given view, and what are the consequences of that view, even if one does not share the intuitions that motivate the view. (Markosian 1998a: 227)

I do not share Markosian's modal intuitions. But I share his sentiments about the value of arguments that employ modal intuitions. Although arguments employing modal intuitions will not persuade those who do not share them, it is worthwhile to see how far views motivated by them can be pushed.

Secondly, since the case for the Brutal View is an argument via elimination, I must acknowledge the possibility that I have failed to consider other possible answers to the Simple Question. My defence is that I am unable to think of what they might be. I would be happy if someone else were to produce a new plausible answer to the Simple Question. As Markosian noted, the Simple Question deserves more attention.[51]

REFERENCES

Armstrong, D. M. (1989) *A Combinatorial Theory of Possibility* (Cambridge: Cambridge University Press).

——(1997) *A World of States of Affairs* (New York: Cambridge University Press).

Barnette, R. L. (1978) ''Does Quantum Mechanics Disprove the Principle of the Identity of Indiscernibles?'', *Philosophy of Science* 45: 466–70

Cartwright, Richard (1975) ''Scattered Objects'', in Keith Lehrer, (ed.), *Analysis and Metaphysics* (Dordrecht, Holland: D. Reidel Publishing Company), 153–71.

Chalmers, David (1996) *The Conscious Mind: In Search of a Fundamental Theory* (Oxford University Press).

Chisholm, Roderick (1973) ''Parts as Essential to Their Wholes'', *Review of Metaphysics* 26: 581–603.

[51] I thank Phillip Bricker, Jake Bridge, Lynne Rudder Baker, Ben Caplan, Cody Gilmore, John Hawthorne, Chris Heathwood, Hud Hudson, Ned Markosian, Carl Matheson, Jonathan Schaffer, Tim Schroeder, Ted Sider, Ryan Wasserman, and Dean Zimmerman for helpful comments on earlier drafts of this chapter.

——(1976) *Person and Object: A Metaphysical Study* (La Salle, Illinois: Open Court Publishing Co.).

Cortens, Andrew (1999) *Global Anti-Realism: A Metaphilosophical Inquiry* (Westview).

Cortes, Alberto (1976) ''Leibniz's Principle of the Identity of Indiscernibles: A False Principle'', *Philosophy of Science* 43: 491–505.

Gendler, Tamar and Hawthorne, John (2002) *Conceivability and Possibility* (Oxford University Press).

Gilmore, Cody (2004) *Material Objects: Metaphysical Issues* (dissertation, Princeton University).

Ginsberg, Allen (1981) ''Quantum Theory and the Identity of Indiscernibles'', *Philosophy of Science* 48: 487–91

Greene, Brian (1999) *The Elegant Universe* (New York, NY: Vintage Books).

Hudson, Hud (2001) *A Materialist Metaphysics of the Human Person* (Ithaca: Cornell University Press).

Jubien, Michael (1993) *Ontology, Modality, and the Fallacy of Reference* (Cambridge: Cambridge University Press).

Lewis, David (1986) *On the Plurality of Worlds* (Oxford: Basil Blackwell).

——(1997a) ''New Work for a Theory of Universals'', in *Papers in Metaphysics and Epistemology* (Cambridge: Cambridge University Press).

——(1997b) ''Humean Supervenience Debugged'', in *Papers in Metaphysics and Epistemology* (Cambridge: Cambridge University Press).

McDaniel, Kris (2003) ''Against MaxCon Simples'', *Australasian Journal of Philosophy* 81: 265–75.

McKinnon, Neil (2003) ''Vague Simples'', *Pacific Philosophical Quarterly* 84: 394–7.

Markosian, Ned (1998a) ''Simples'', *Australasian Journal of Philosophy* 76: 213–26.

——(1998b) ''Brutal Composition'', *Philosophical Studies* 92: 211–49.

——(2004) ''SoC it to Me? Reply to McDaniel on MaxCon Simples'', *Australasian Journal of Philosophy* 82: 332–40.

——(2004) ''Simples, Stuff, and Simple People'', *The Monist* 87: 405–28.

Maudlin, Tim (1998) ''Part and Whole in Quantum Mechanics'', *Interpreting Bodies: Classical and Quantum Objects in Modern Physics*, Elena Castellani (ed.) (Princeton University Press: Princeton), 46–60.

Merricks, Trenton (1998) ''On Whether Being Conscious is Intrinsic'', *Mind* 107: 845–6.

——(1999) ''Composition as Identity, Mereological Essentialism, and Counterpart Theory'', *Australasian Journal of Philosophy* 77: 192–5.

——(2003) ''Maximality and Consciousness'', *Philosophy and Phenomenological Research* 66: 150–8.

Parsons, Josh (2000) ''Must a Four-Dimensionalist Believe in Temporal Parts?'', *The Monist* 83: 399–418.
Rosen, Gideon, and Dorr, Cian (2003) ''Composition as a Fiction'', in *The Blackwell Companion to Metaphysics*, ed. Richard Gale (Blackwell).
Scala, Mark (2002) ''Homogeneous Simples'', *Philosophy and Phenomenological Research* 64: 393–7.
Schaffer, Jonathan (2003a) ''The Problem of Free Mass: Must Properties Cluster?'', *Philosophy and Phenomenological Research* 66: 125–38.
——(2003b) ''Is There a Fundamental Level?'' *Nous* 37: 498–517.
——(forthcoming) ''Monism''.
Sidelle, Alan (1989) *Necessity, Essence, and Individuation: A Defense of Conventionalism* (Ithaca: Cornell University Press).
Sider, Theodore (1993) ''Van Inwagen and the Possibility of Gunk'', *Analysis* 53: 285–9.
——(1995) ''Sparseness, Immanence, and Naturalness'' *Nous* 29: 360–77.
——(1997) ''Four-Dimensionalism'', *Philosophical Review* 106: 197–231.
——(2001) *Four-Dimensionalism: An Ontology of Persistence and Time* (Oxford: Clarendon Press).
——(2003) ''Maximality and Microphysical Supervenience'', *Philosophy and Phenomenological Research* 66: 139–49.
——(forthcoming) ''Parthood'', *Philosophical Review*.
Simons, Peter (1987) *Parts: A Study in Ontology* (Oxford: Oxford University Press).
——(1994) ''Particulars in Particular Clothing: Three Trope Theories of Substance'', *Philosophy and Phenomenological Research* 54(3): 553–75.
Teller, Paul (1983) ''Quantum Physics, the Identity of Indiscernibles and Some Unanswered Questions,'' *Philosophy of Science* 50: 309–19.
van Inwagen, Peter (1981) ''The Doctrine of Arbitrary Undetached Parts'', *Pacific Philosophical Quarterly* 62: 123–37.
——(1990) *Material Beings* (Ithaca: Cornell University Press).
——(1994) ''Composition as Identity'', in J. E. Tomberlin (ed.), *Philosophical Perspectives 8, Logic and Language* (Atascadero, CA: Ridgeview), 207–20.
Wasserman, Ryan (2003). ''Review Article: Hud Hudson's *A Materialist Metaphysics of the Human Person*'', *Philo* 6: 307–13.
Zimmerman, Dean W. (1995) ''Theories of Masses and Problems of Constitution'', *Philosophical Review* 104: 53–110.
——(1997) ''Indivisible Parts and Extended Objects: Some Philosophical Episodes from Topology's Prehistory'', *The Monist* 79: 148–80.
——(1998) ''Temporary Intrinsics and Presentism'', in Zimmerman and Peter van Inwagen (eds.), *Metaphysics: The Big Questions* (Cambridge, MA: Blackwell).

——(2000) ''Could Extended Objects Be Made Out of Simple Parts? An Argument for Atomless Gunk'', *Philosophy and Phenomenological Research* 56: 1–29.

——(2002) ''Scala and the Spinning Spheres'', *Philosophy and Phenomenological Research* 64: 398–405.

This page intentionally left blank

9. Contact and Continuity

William Kilborn

A BRIEF SKETCH

This chapter develops a new approach to the Brentanian puzzle of contact and continuity. Various versions of the puzzle or closely related considerations thereof have received discussion in the works of Bolzano, Brentano, Leonardo, and Peirce, and more recently in Casati and Varzi's *Parts and Places*, in Hawthorne's ''Before-Effect and Zeno Causality'', in Hudson's *The Metaphysics of Hyperspace*, in Kline and Matheson's ''The Logical Impossibility of Collision'', in Varzi's ''Boundaries, Continuity, and Contact'', and in Zimmerman's ''Could Extended Objects Be Made Out of Simple Parts? An Argument for 'Atomless Gunk' ''. In his ''Simply Possible'', Theodore Sider has proposed a prima facie plausible solution to the Brentanian puzzle. In the present work, I point out that there are other intriguing puzzles of contact and continuity. An interesting result is that Sider's proposal cannot help solve every puzzle of contact. It seems that this undercuts much, if not all of the motivation for endorsing Sider's proposal stemming from certain considerations regarding the original Brentanian puzzle. Another interesting result is that a solution championed by some earlier authors to the Brentanian puzzle appears now to be quite plausible as the correct solution to every puzzle of contact. The proposed solution is that it is metaphysically impossible for space to be continuous. But, as it will become clear, there are important costs that come with endorsing any response to the puzzles of contact.

§1 SOME FAMILIAR TERMS OF THE CONTROVERSY

Some terminology is needed to lay the groundwork for the following discussion. In the following stipulations, *p* and *R* are variables ranging

over spatial regions, S is a variable ranging over regions *simpliciter*, x is a variable ranging over objects, and any relevant temporal qualifications have been left out:

(C0) 'S is continuous' $=_{\text{df}}$ every extended spatial subregion of S has continuum many points as subregions, S has extended spatial subregions, and S decomposes without remainder into points.

(C1) 'the complement of R' $=_{\text{df}}$ the region which has as subregions every point which is not a subregion of R.

(C2) 'R is an open sphere about p' $=_{\text{df}}$ R has as subregions all those points that are less than some fixed distance from p, R has no other points as subregions, and p is a point.

(C3) 'p is a boundary point of R' $=_{\text{df}}$ every open sphere about p has as subregions a subregion of R and a subregion of the complement of R, and p is a point.

(C4) 'the closure of R' $=_{\text{df}}$ the region which has as subregions every point which is a subregion of R and every boundary point of R, but has no other points as subregions.

(C5) 'R and R^* are separated' $=_{\text{df}}$ no subregion of R is also a subregion of the closure of R^*, and no subregion of R^* is also a subregion of the closure of R.

(C6) 'R is disconnected' $=_{\text{df}}$ there are two separated subregions of R such that every point which is a subregion of R is a subregion of one of the two separated subregions.

(C7) 'R is connected' $=_{\text{df}}$ R is not disconnected.

(C8) 'x is an open object' $=_{\text{df}}$ x is exactly located at a region that has none of its boundary points as subregions.[1]

(C9) 'x is a closed object' $=_{\text{df}}$ x is exactly located at a region that has all of its boundary points as subregions.

(C10) 'x is a partially open object' $=_{\text{df}}$ x is exactly located at a region that has some but not all of its boundary points as subregions.

(C1) through to (C10) are borrowed from Richard Cartwright's 'Scattered Objects',[2] but feature some important changes here. It is important to note that we would need a counterfactual or other reading of some of the above definitions when describing certain odd-shaped

[1] I assume that, like material objects, regions could be located at regions, but unlike material objects, they cannot be occupants of regions.

[2] Cartwright (1987).

spacetimes if we do not want to trivialize or mistakenly use them.[3] Also important to note for the following discussion is that if spacetime is essentially continuous and material objects cannot be co-located, then, necessarily, for any distinct objects, x and y, x and y are in perfect contact at some time, t, *iff* there is a region, R, such that (i) every point-sized subregion of R is either a subregion of the region at which x at t is exactly located or a subregion of the region at which y at t is exactly located, (ii) some point which is a subregion of the region at which x at t is exactly located and some point which is a subregion of the region at which y at t is exactly located are both subregions of R, and (iii) R is connected.[4]

§2 THE ORIGINAL PUZZLE: CONTACT, CONTINUITY, AND CO-LOCATION

Suppose, possibly, spacetime is continuous. Then, there could be material objects which are either open, closed, or partially open. The following is a possible situation: a material object, O_1, is rapidly approached by a material object, O_2, on a collision course with O_1. For ease of exposition, let the material objects be solid spheres. If one of the objects were open and the other closed, they would, unless there were some good reason for their being kept apart, come into perfect contact.

[3] For example, suppose there is a world with a closed cubical region, Cube, and an open region, OR, completely surrounding it so that there is no region of space between them. Let 'Point' name a boundary point of Cube. Suppose also that there is another world much like the one just described, except that OR doesn't exist, and no other region in any way fills in for it. In this second world, Point, or a relevantly similar counterpart, is an outermost point of Cube, or a relevantly similar counterpart, just as in the first world. In accordance with the Lewisian conception of distance relations and the strict reading of the definitions, Point is not a boundary point of Cube in the second world since there are some open spheres about it which don't have the complement of Cube as a subregion. However, it seems very natural to infer from the description of the second world that Point is a boundary point of Cube in the second world. For a sophisticated discussion of various views regarding distance relations, see Phillip Bricker's "The Fabric of Space: Intrinsic *vs.* Extrinsic Distance Relations".

[4] Some further notes: it may seem that the possible topology of the temporal dimension is not at issue in the following discussion. However, it (at least prima facie) seems that if, possibly, time is continuous, then, possibly, space is continuous. It will become clear that, for this reason, the following discussion seems pertinent to the possible topology of time as well as to the possible topology of space. Also, 'two material objects are separated at t' is shorthand for stating that the two regions which the material objects at t exactly occupy are separated, and the strength of the modal qualifications throughout the chapter, unless otherwise specified or made obvious by the context, are metaphysical.

Suppose instead, however, that the objects are both open. Without experiencing a change in shape and without having parts co-located, the objects would always be separated by at least a single point in space. Suppose, though, that O_1 and O_2 are both closed. Then, without changing in shape and without having parts co-located, the objects would always be separated by a nonzero distance. Suppose, finally, that at least one of the objects is partially open. Then, there would still be a path in spacetime where though O_2 is on a collision course with O_1, without either object changing in shape, O_2 cannot come into perfect contact with O_1.[5]

Given the *de dicto* claim that spacetime could be continuous, it is possible that a material object is on a collision course with another material object and that both objects are open or closed, or at least one of them is partially open. But then, in uncountably many possible situations where material objects are on a collision course with each other, the objects cannot help but always be separated by at least a single spatial point, or even some nonzero distance, without either object changing in shape and without interpenetrating. We may reflect upon what happens in such situations.

In some situations, one of the material objects involved undergoes a change in shape so that the two come into perfect contact with each other. Physical laws or forces may in part be responsible for a rearrangement or an annihilation of parts of the objects. In some situations, laws or forces may in part be responsible for one of the objects moving so as to avoid the other when they get very close. In other situations, objects may suddenly appear so that perfect contact is made between several objects.[6] *And so on.*

Thus, given that spacetime could be continuous, it is surely a contingent matter as to whether or not there are laws, forces, or whatever which are responsible for changes in shape, sudden appearances or disappearances of material objects, and other disturbances. But, moreover,

[5] The path would leave O_1 and O_2 at least a nonzero distance apart or at least a single spatial point apart from each other. By 'x collides with y at t' I just mean that x and y are in perfect contact at t, and arbitrarily close but prior to t, x and y are not in perfect contact. In light of this, 'collision course' may appear to be a somewhat misleading term. Also, I assume any colocation would preclude perfect contact.

[6] For example, the appearance of a single point-sized material object is enough to close the gap between two spheres which are open. Please see Hawthorne (2000) and Hudson (2005: ch. 3 §6) for more discussion about what could happen in these situations assuming that spacetime could be continuous.

it is quite plausible that they are not present in every previously described possible collision course situation.[7] Therefore, under the supposition that spacetime could be continuous, it seems that there are at least some possible worlds in which

(i) a material object is approached by another material object on a collision course with it,
(ii) the objects are both open, both closed, or at least one of the objects is partially open,
(iii) if either object is partially open, then unless they interpenetrate, the approaching object follows a path in spacetime which leaves the objects so that they cannot, given their topological structures, help but be spatially separated, and
(iv) there are no laws, forces, or other contingent matters[8] which in any way disturb the approaching object during its voyage (or any of the closures of the spatial regions that it comes to exactly (spatially) occupy), or which in any way disturb the object being approached (or any of the closures of the spatial regions that it comes to exactly occupy).[9]

It may seem at this point that if there are any possible worlds where (i) through to (iv) obtain, the objects in question cannot come into perfect contact with each other, and thus, they cannot help but remain spatially separated from each other even though one is approached by another on a collision course with it. But, given (iv), it seems that there would not be an adequate explanation of what would happen so that the objects remain spatially separated from each other. At least the properties of,

[7] For similar lines of thought, see Hudson (2005: ch. 3 §6) and Zimmerman (1996a: 1–15).

[8] I also mean to include brute matters of fact. By '*x* is a brute matter of fact' I mean *x* is a contingently true proposition (or otherwise a contingent state of affairs) which is such that there is no sufficient reason for its being true (or obtaining). Some philosophers think that the existence of brute matters is established by its being the only genuinely plausible response to one of the most fascinating puzzles in metaphysics. Please see Peter van Inwagen's *An Essay on Free Will*, 202–4 for more details.

[9] Let us suppose the closures remain undisturbed except perhaps by being occupied by either of our two material objects.

and the relations between, the material objects would not provide an adequate explanation.[10] Thus, it may seem that even if spacetime could be continuous, there are no worlds in which (i) through to (iv) obtain.

[10] Joshua Spencer and Chris Tillman have suggested that, in worlds where conditions (i) through to (iv) obtain, just the topological structures of two material objects in question would account for (i.e. be that wholly in virtue of which) whatever happens so that they remain spatially separated from each other. Suppose that, in such a world, an open material object approaches a second open material object. Any region in perfect contact with the second object would at best be partially open. Certainly, the difference in topology between a material object and a spatial region is metaphysically sufficient for the material object's not exactly (spatially) occupying the region. Thus, the difference in topology (at all times) alone is metaphysically sufficient for the objects not coming into perfect contact. Still, if conditions (i) through to (iv) were to obtain, without laws, forces, or other such contingent matters explaining a material object's swerving, slowing down, or whatever on approach toward another material object, it seems the mere shapes of the material objects or regions involved would not provide a complete explanation for *what happens* so that the two objects stay spatially separated. It may be suggested that, owing to their shapes, an open object's approach would be abruptly affected just a single spatial point apart from a second open object in some of the worlds in which (i) through to (iv) obtain. However, there is no equally elegant suggestion as to what would happen if the objects were both closed (or, under certain conditions, if one were partially open and the other were closed or partially open) since the objects must remain a nonzero distance apart from each other. Given Spencer and Tillman's suggestion, the claim that the approaching object would be disturbed at a given nonzero distance apart from the other object seems mistaken since the mere topology of the objects would not explain why the approaching object would be disturbed at some nonzero distance from the other object, rather than a smaller nonzero distance. Therefore, given that there are worlds in which conditions (i) through to (iv) obtain, it seems that just the shapes of the objects (and regions involved) would fail to account for what would happen in every such world.

In response, a proponent of Spencer and Tillman's suggestion may endorse the O-Theory (so dubbed in Hudson 2005: 48) and take it that perfect contact between two material objects is a matter of being separated by one spatial point (e.g. see Cartwright 1987: 172). The O-Theory is the view that, necessarily, any material object is open. The O-Theory seems false since it is a fact that, necessarily, if space is n-dimensional at t, then at t, any open object is n-dimensional. Thus, according to the O-Theory, if space were suddenly to gain a dimension, then all material objects would go out of existence without first changing in shape. However, this seems false since it seems not to be a property of every material object that such a change in dimension would either prompt a corresponding change in the dimensionality of it or force it to go out of existence (through conversation, I have found that Hud Hudson has independently come up with the same argument). Of course, there are many other good reasons for thinking that it is possible (and even actual) that there is a *less-than-n*-dimensional material object in an n-dimensional space, and these reasons are at odds with the O-Theory. Reasons for rejecting the O-Theory aside, there is another good reason why a proponent of Spencer and Tillman's suggestion should not endorse the O-Theory: given that spacetime could be continuous, regions right up against a perforation in spacetime could come in open, closed, and partially open varieties. It will become clear that, as a consequence, the O-Theorist must either face the trilemma which I discuss in the next section, or take it that (i) in some worlds, a material object approaching a perforation does not remain spatially separated from the perforation, (ii) in other worlds, a material

However, it is still quite plausible that, necessarily, if spacetime is continuous, then there are possible worlds in which (i) through to (iv) obtain. After all, the existence of such interfering laws, forces, or whatever would be merely a contingent matter. By a simple, plausible argument from subtraction, we may infer that if spacetime could be continuous, then it is not a metaphysically necessary condition of there being continuous spacetime that any such laws, forces, or whatever are present in every previously described collision course situation.[11]

Thus, there may seem to be a tension between the two following claims,

(Continuity) Possibly, spacetime is continuous.
(No Laws) Necessarily, if spacetime is continuous, then, possibly, (i) through to (iv) obtain.

The argument above may be framed as,

(1) (Continuity) ⊃ ∼(No Laws).
(2) ∼∼(No Laws). /∴ ∼(Continuity).[12]

What's puzzling here is that it is at least prima facie plausible that the conclusion is false even though each of the premises are also at least prima facie plausible.

One response to this argument is that premise (1) is false. Instead of being forced to choose between rejecting (Continuity) and rejecting (No Laws), as the response goes, the puzzle forces us to admit the possibility of co-located material objects. This possibility would eliminate the difficulties presented here for (Continuity) since there would

object approaching a perforation seems condemned to be always separated from the perforation by a single spatial point, and (iii) whatever it is that happens in these worlds is completely explained by the shapes of the material object and the perforated region. However, choosing the former option seems to undermine the apparent main reason for becoming an O-Theorist in this instance, while choosing the latter option just seems extravagant. For discussion of some reasons in support of, and against, the O-Theory and its rivals, see Hudson (2005: ch. 2), Zimmerman (1996a), and Zimmerman (1996b).

[11] Subtraction considerations aren't an essential part of any of my arguments. For an example of an argument from subtraction (but for a different conclusion), Rodriguez-Pereyra (1997).

[12] There are some similarities between the argument presented here and one found in Zimmerman (1996a: 1–15). There are, however, many important differences. For example, Zimmerman does not insist that the existence of continuous spacetime is impossible. In fact, Zimmerman is quite fond of the O-Theory (see note 10) and defends it from attack: see Zimmerman (1996a: 17–19) and Zimmerman (1996b: 160–5). What's not possible, according to Zimmerman, is that there are extended occupants which decompose without remainder into point-sized simples. In setting up the puzzle in this section, I have also found helpful an excellent discussion of the Brentanian puzzle presented in Hudson (2005: ch. 3 §6).

be worlds in which (i) through to (iv) obtain, and in such worlds, material objects would not remain spatially separated from each other but instead, would go through each other.[13] Thus, it seems that either (Continuity) is false, (No Laws) is false, or (Co-location) is true.

(Co-location) Possibly, there are distinct material objects, *x* and *y*, some time, *t*, and a region of continuous spacetime, *R*, such that *x* at *t* and *y* at *t* each exactly occupy *R*, and *x* at *t* and *y* at *t* do not share parts.

Accepting (Co-location) is thought to be the correct solution to this puzzle in Theodore Sider's ''Simply Possible''.[14] It is worth noting that (Co-location) is a highly controversial claim, and many philosophers have denied it.

I now introduce a thought experiment from which it seems we learn that either (Continuity) is false or a principle very similar to (No Laws) is false, and the truth of (Co-location) cannot resolve this tension. An interesting virtue of the following puzzle is that, unlike the original puzzle of contact, only a single material object is required to generate it.[15]

§3 CONTACT, CONTINUITY, AND ODD-SHAPED SPACETIMES

According to philosophical orthodoxy at this time, spacetime could have different metrical and topological properties from those that it has.[16]

[13] That is, the approaching objects would not deviate from their courses at any time prior to being colocated.

[14] Sider (2000). Sider doesn't state that we are presented with a good argument for (Colocation) from these considerations. Rather, any good reason for thinking that (Colocation) is true is presumably had on independent grounds. On an important, related note, it seems that a bare principle which states that being material entails being not colocated, or which states that being material entails being possibly colocated, is not analytically true or false. There are clearly some doubts about this, however (e.g. see Zimmerman 1996a: 3). For my part, I endorse the Spatial Occupancy Account of the concept of a material object which states, necessarily, for any object, *x*, *x* is a material object *iff* *x* is an exact (spatial) occupant of a spatial region. To see an argument for a very similar account, see Ned Markosian's ''What Are Physical Objects?'' If this seems mistaken, let us take this definition stipulatively. On this definition of 'material object', it seems that any such bare principle is not analytically true or false.

[15] Please see the end of note 10 for one reason why.

[16] Please see Bricker (1991) and Huggett (2006) for related discussion on the matter.

A specific suggestion in the literature is that spacetime could be perforated.[17] *Being perforated* is a topological property that spacetime does not seem to exemplify but seemingly could exemplify. If spacetime could be perforated, then the following is a possible situation: a perforation (speaking loosely) is rapidly approached by a material object on a collision course with it.[18] Moreover, it is quite plausible that if spacetime could be continuous and if spacetime could be perforated, then spacetime's being continuous does not entail that it is not perforated. Importantly, then, if spacetime could be continuous, there could be a perforation which is either 'open', 'closed', or 'partially open'.

A brief explanation seems worthwhile. Suppose there are three instantaneous material objects which are spatially separated from each other and do not share any parts. One of them is open, another is closed, and another is partially open. Suppose further that, at some later time, all and only the spatial points which the objects wholly occupied are annihilated. As a consequence, each of the three resulting perforations would have a different shape.

Again, to illustrate, suppose that an open material object is traveling on a course to exactly (spatially) occupy an open spatial region, *R*. Suppose also that, before the object reaches *R*, all and only the spatial points which are subregions of *R* are annihilated. The open object would then be approaching an 'open' perforation. Moreover, the object could not help but always be separated from the perforation by at least a single spatial point without either it or certain super-regions at which it is located changing in shape, whereas there would not be any such limitation on how close it could get if it were to approach a 'closed' perforation. This is due in part to the fact that any region which had been in perfect contact with *R* could not have been open but would at best have been partially open, and an open object cannot exactly occupy any spatial region which is not open without first either the object or the region changing in shape. An intuitive lesson we may draw is this: if spacetime could be both continuous and perforated, then in some worlds, a material object approaching a perforation cannot help but always have a region of space between it and the perforation.

[17] Bricker (1993: 275–6) and Huggett (2003: 283). Note that Nick Huggett is a relationalist about spacetime.

[18] Talk of there being a 'spatiotemporal perforation' is a mere *façon de parler* since there cannot be any genuine entity which falls under the concept. Also, talk of 'open', 'closed', and 'partially open' non-receptacles may be given a counterfactual gloss.

However, assuming that a continuous, perforated spacetime is possible, it seems that spacetime's having these features does not entail the existence of interfering laws, forces, or other contingent matters in situations very similar to those described in the previous puzzle. Much as with the previous puzzle, then, it seems that if spacetime could be continuous and perforated, there are some possible worlds that have (i)* through to (iv)* as features of them:

(i)* a material object is on a collision course with a perforation,

(ii)* either both the material object and the perforation are open or closed, or at least one of them is partially open,

(iii)* if either the material object or perforation is partially open, then the material object follows a path in spacetime which leaves the object so that it cannot help but be spatially separated from the perforation, and

(iv)* there are no laws, forces, or other contingent matters which in any way disturb the approaching object during its voyage (or any of the closures of the spatial regions that it comes to exactly (spatially) occupy), or which in any way disturb the nearby perforated regions being approached.

In every world which has (i)* through to (iv)* as features, a material object on a collision course with a perforation must remain spatially separated from the perforation. But, given (iv)*, it seems there would not be an adequate explanation of what would happen so that the object remains spatially separated. Thus, it seems that even if spacetime could be continuous and perforated, there are no possible worlds which have (i)* through to (iv)* as features. Moreover, Sider's solution to the previous puzzle is clearly no help in alleviating a move toward affirming the antecedent or denying the consequent of the premise below. No principle of co-located entities could come to the rescue since a material object cannot be located at a non-receptacle.[19] The support for the following premise (1)* is clear.

(C-Perforated) Possibly, spacetime is continuous and perforated.
(No Laws)* Necessarily, if spacetime is continuous and perforated, then, possibly, (i)* through to (iv)* obtain.

(1)* (C-Perforated) $\supset \sim$(No Laws)*.

[19] This easily follows from the Spatial Occupancy Account of material objects (see note 14), but there are also other good reasons for thinking it is true.

As with premise (1), the puzzle behind (1)* is that both the antecedent and the denial of the consequent are at least prima facie plausible. It seems that we are forced either to reject that spacetime could be perforated, reject (No Laws)*, or reject (Continuity).[20]

As for the first option, I take it that spacetime could be perforated since it is quite problematic to deny this. For example, it seems that spacetime in fact changes in shape when interacting with at least some material objects. But, it seems there is no necessary prohibition on the shapes of spacetime which we are interested in. Such a restriction would be too arbitrary or just plain strange. After all, all that we would need is for spacetime to stretch or tear a certain way.[21, 22]

With regard to rejecting (No Laws)*, as with denying principle (No Laws) of the original Brentanian puzzle, there is good reason for thinking that it is a mistake. It seems that rejecting (No Laws)* is about as problematic as denying (No Laws) owing to the similar intuition that it would not be a metaphysically necessary condition of spacetime's being both continuous and perforated that interfering laws, forces, or other contingent matters are present in every previously described collision course situation involving perforations.[23] After all, there is a simple subtraction argument available for (No Laws)* which is quite similar in

[20] Being perforated, however, is not the only property that gives rise to such puzzles. For example, it is quite easy to see that the properties of *being a finite volume* and *being the shape of an infinitely long cylindrical region* each do this as well. (It may seem that these puzzles are only motivated by the possibility of space with an edge. This may not be correct, however. For example, if there could be regions which are impenetrable by material objects on approach due to their geometrical properties), then we may have a similar puzzle.) From this, it seems that we are forced either to *reject that space could have properties like these*, reject (No Laws)*, or reject (Continuity). To simplify matters, I focus my discussion on the property of being perforated and the above puzzle, but what I say goes *mutatis mutandis* for some of the other properties and the similar puzzles behind them.

[21] This is, however, not the only good reason for thinking that rejecting the possibility of perforated spacetimes is a mistake. For example, some other reasons are discussed in Bricker (1991) and Bricker (1993).

[22] Here is how stretching would help: Suppose, as many of the folk suspect, space is infinitely large, Euclidean, and three-dimensional, say at t_0. Suppose also that a region of space is stretched at t_1 to produce a four-dimensional spatial region which itself is only finitely large. At t_1, there would be an edge to space which a material object could travel to. Another way in which stretching would help is if a region in such a space were stretched into a loop, thereby creating a perforated region.

[23] Perhaps such necessitation of interfering laws pertaining to parts of spacetime is much more problematic than the previously described necessitation of interfering laws pertaining just to material objects.

force to, and just as plausible as, the one in favor of (No Laws). For this reason, it seems that the denial of (No Laws)* is implausible. From these considerations, it seems that the argument against the possibility of continuous spacetime can be put quite strongly.

For some final thoughts, suppose (Co-location) is endorsed as a solution to the Brentanian puzzle for the sole reason that both (Continuity) may be saved and a commitment to ∼(No Laws) may be avoided. After all, rejecting either (Continuity) or (No Laws), it may seem, is too extravagant a move. There are excellent reasons for thinking that if (Continuity) is true, then (C-Perforated) is true. Moreover, it is quite clear that premise (1)* is true. Given (continuity), then, we have an excellent reason for thinking that (No Laws)* is false. However, it seems that rejecting (No Laws)* is about as problematic as denying (No Laws). But, then, without independent reasons in favor of (Co-location), much or all of the motivation for endorsing it as a solution to the first puzzle seems undercut.

If Sider is right, though, and (Continuity) is true, then either spacetime cannot be perforated, spacetime could be continuous and could be perforated but could not be both continuous and perforated or there would be interfering laws, forces, or other contingent matters in every previously described possible collision course situation involving perforations. There are some good reasons for thinking that the consequent of this conditional is quite implausible, and there are attractive alternatives which don't demand a similar, implausible choice. Under the scope of these considerations alone, it seems that rejecting (Continuity) to avoid having to accept any one of these options is the most plausible solution to the puzzles of contact and continuity.[24, 25]

[24] Let 'x is a material simple' $=_{\mathrm{df}}$ x is a material object, and x has no proper parts. The above result may spell trouble for a natural view regarding material simples, namely the Pointy View, which states, necessarily, for any object, x, x is a material simple *iff* x exactly occupies a spatiotemporal point (the name, which is given to roughly the same view, comes from Markosian 1998: 216). This is since it may seem that (as many philosophers seem to think) if continuous spacetime were impossible, then there could not be spatiotemporal points, but it also seems there could be material simples. On the other hand, it seems to be quite a cost of denying (Continuity) to also have to deny the Pointy View or deny that there could be material simples. This appears to be just one of many important costs of denying (Continuity).

[25] Thanks to Hud Hudson, Christian Lee, Chris Tillman, Stephan Torre, and Dean Zimmerman for helpful comments on an earlier draft of this chapter. Special thanks to Joshua T. Spencer for his insightful and thorough comments on several drafts.

REFERENCES

Bolzano, Bernard (1950) *Paradoxes of the Infinite*, Fr. Přihonský, trans. (London: Routledge).

Brentano, Franz (1988) *Philosophical Investigations on Space, Time and the Continuum*, Barry Smith, trans. (London: Croom Helm).

Bricker, Phillip (1991) ''Plenitude of Possible Structures'', *The Journal of Philosophy* 88: 607–19.

——(1993) ''The Fabric of Space: Intrinsic *vs*. Extrinsic Distance Relations'', *Midwest Studies in Philosophy*, vol. 18, Peter A. French, Theodore E. Uehling, Jr., and Howard K. Wettstein (eds.) (Notre Dame, IN: University of Notre Dame Press), 271–94.

Cartwright, Richard (1987) ''Scattered Objects'', *Philosophical Essays* (Cambridge, MA: The MIT Press), 171–86.

Casati, Roberto, and Varzi, Achille C. (1999) *Parts and Places: The Structures of Spatial Representation* (Cambridge, MA: The MIT Press).

Hawthorne, John (2000) ''Before-Effect and Zeno Causality'', *Noûs* 34: 622–33.

Hudson, Hud (2005) *The Metaphysics of Hyperspace* (Oxford: Clarendon Press).

Huggett, Nick (2003) ''Mirror Symmetry: What is it for a Relational Space to be Orientable?'', *Symmetries in Physics: Philosophical Reflections*, Katherine Brading and Elena Castellani (eds.) (Cambridge: Cambridge University Press), 281–8.

——(2006) ''The Regularity Account of Relational Spacetime'', *Mind* 115: 41–73.

Kline, A. David, and Matheson, Carl A. (1987) ''The Logical Impossibility of Collision'', *Philosophy* 62: 509–15.

Leonardo da Vinci (1938) *The Notebooks of Leonardo da Vinci*, Edward MacCurdy (ed.), trans. (London: Reynal).

Markosian, Ned (2000) ''What Are Physical Objects?'', *Philosophy and Phenomenological Research* 61: 375–95.

—— (1998) ''Simples'', *Australasian Journal of Philosophy* 76: 213–28.

Peirce, Charles Sanders (1933) ''The Logic of Quantity'', *Collected Papers of Charles Sanders Peirce*, vol. 4, Charles Hartshorne and Paul Weiss, (eds.) (Cambridge, MA: Harvard University Press), 59–131.

Rodriguez-Pereyra, Gonzalo (1997) ''There Might be Nothing: The Subtraction Argument Improved'', *Analysis* 57: 159–66.

Sider, Theodore (2000) ''Simply Possible'', *Philosophy and Phenomenological Research* 60: 585–90.

van Inwagen, Peter (1983) *An Essay on Free Will* (Oxford: Clarendon Press).

Varzi, Achille C. (1997) ''Boundaries, Continuity, and Contact'', *Noûs* 31: 26–58.

Zimmerman, Dean W. (1996a) ''Could Extended Objects Be Made Out of Simple Parts? An Argument for 'Atomless Gunk' '', *Philosophy and Phenomenological Research* 56: 1–29.

——(1996b) ''Indivisible Parts and Extended Objects: Some Philosophical Episodes from Topology's Prehistory'', *The Monist* 79: 148–80.

Part V

FREE WILL

This page intentionally left blank

10. Metaphysical Compatibilism's Appropriation of Frankfurt

Ted A. Warfield

I. TWO DEBATES AND THE FREEDOM/RESPONSIBILITY ARGUMENT

In discussions of causal determinism, the names ''Compatibilist'' and ''Incompatibilist'' are sometimes used in different ways. When focusing on the question of whether causal determinism precludes *metaphysical freedom* (or ''free will'' as it is unfortunately often called) a ''compatibilist'' is one who believes that the truth of causal determinism is consistent with the existence of human freedom and an ''incompatibilist'' is one who believes that these theses are not consistent. These same labels are also used, in the obvious parallel way, in discussions of whether causal determinism precludes *moral responsibility*.

The freedom/determinism debate and the moral responsibility/determinism debate are, of course, tied together by more than this parallel use of terminology. Most philosophers who work on the metaphysics of freedom at least partly motivate this work by gesturing towards issues about moral responsibility. This is, for example, what is commonly included in the advertising copy for most books about the metaphysics of freedom in a rather transparent attempt to draw in a wider audience. An additional connection between these two debates is that most philosophers are compatibilists about both issues.[1] These are only three examples of connections between these debates. The full extent of this connection is surely tighter than the combination of terminological

[1] I do not think that this claim is true of most philosophers who work professionally on these issues. I think that most of these philosophers are incompatibilists about causal determinism and human freedom and I think these philosophers are evenly divided on the causal determinism/moral responsibility compatibility issue.

overlap, advertising copy, and close percentages of philosophers adopting eponymous positions suggests. Indeed, most philosophers concerned with this pair of compatibility issues believe that the two debates are closely connected in some *philosophically* important way because they believe that, at a minimum, there is some at least fairly intimate connection between freedom and moral responsibility.

I agree with the widely shared view that these two debates are intimately connected because metaphysical freedom and moral responsibility are in some way importantly connected. I will identify one initially plausible view of the nature of the connection between metaphysical freedom and moral responsibility. This view will suggest an argument that I believe is widely implicitly endorsed though never, to my knowledge, explicitly formulated and defended. The argument, if sound, forces philosophers to adopt compatibilism about causal determinism's relation to both moral responsibility and metaphysical freedom. Furthermore the considerations driving the argument (even if the argument is not sound) suggest that one cannot mix and match verdicts on these two compatibility debates. That is, the considerations suggest that one must be either a compatibilist about both issues or an incompatibilist about both issues.

It seems likely to me that a fair number of philosophers have been mistakenly led to these conclusions by the argument I will be rejecting.[2] So, if I am right, the mistaken argument to be identified shortly leads to nontrivial philosophical confusions. If by chance I am wrong and what I am so far calling only the ''mistaken argument'' is sound, this has important consequences for some important philosophical debates. Let's move in for a closer look at some of these issues and to identify this argument.

It will be useful to have some stipulative terminology in place so that we do not confuse distinct philosophical positions. As noted above, the two determinism debates share some names for key positions. In an attempt to

[2] In saying that philosophers are ''mistakenly led'' to these compatibilist positions, I mean two things. First, and this is the ''primary meaning'' of the statement for purposes of this chapter, I mean that this particular argument should not lead anyone to any compatibilist position even if that position is true. Secondly, though I will not take the time to argue for this position here, I mean that the argument should not lead anyone to any compatibilist position about causal determinism and metaphysical freedom and/or moral responsibility because these positions are false (see Warfield 1996, 2000, and forthcoming for arguments for this latter claim).

avoid confusion, let's call those who accept compatibilism about causal determinism and responsibility ''R-Compatibilists'' (for ''Responsibility compatibilists'') and those who accept compatibilism about causal determinism and metaphysical freedom ''M-Compatibilists'' (for ''Metaphysical compatibilists''). Use ''R-Incompatibilist'' and ''M-Incompatibilist'' in the obvious parallel manner. Let's call those who accept *both* R- and M-Compatibilism ''Full Compatibilists'' and those who accept *both* R- and M-Incompatibilism ''Full Incompatibilists''. Lastly, let's call those who accept the conjunction of R-Compatibilism with M-Incompatibilism ''Semi-Compatibilists''.[3]

Harry Frankfurt's important work on the Principle of Alternative Possibilities is at the root of the contemporary Full Compatibilist position.[4] My defense of this claim and rational reconstruction of the roots of contemporary Full Compatibilism begins now. Most R-Compatibilists have appropriated Frankfurt's work on PAP in defending R-Compatibilism. Additionally many (perhaps most) R-Compatibilists think that Frankfurt's work on PAP also has important consequences for the defense of M-Compatibilism.

It is not hard to see why philosophers might think that Frankfurt's work on PAP has something to do with the R-Compatibilism debate. Though they are not equivalent notions, ''causal determinism'' and ''lack of alternative possibilities'' are conceptual cousins (it seems that the former entails but is not entailed by the latter). Why, though, would anyone think that Frankfurt's work on moral responsibility (rather than his independent work on *metaphysical freedom*) has important ramifications for the M-Compatibilism debate? The answer to this

[3] Fortunately we will not need to consider the position that combines R-Incompatibilism with M-Compatibilism and so we can avoid introducing the term ''semi-incompatibilist'' into the discussion. Those familiar with John Martin Fischer's important work on these issues know that Fischer has described and defended Semi-Compatibilism (see his 1994 and Fischer and Ravizza 1998). Some who accept M-Incompatibilism think that though moral responsibility is consistent with a lack of alternative possibilities it is not consistent with causal determinism. This position is not ''Semi-Compatibilism'' in my sense: this position is one version of Full Incompatibilism.

[4] As formulated by Frankfurt, PAP says ''an agent is morally responsible for what he has done only if he could have done otherwise'' (Frankfurt 1969: 828). I do not know if Frankfurt would endorse the Full Compatibilist position. Frankfurt's account of freedom does seem to commit him to M-Compatibilism and many people take Frankfurt to be a R-Compatibilist as well. So far as I am aware (but I am no Frankfurt scholar), though Frankfurt has argued that moral responsibility does not require alternative possibilities he has not in print argued or claimed that it is consistent with causal determinism. As indicated in the previous note, this is potentially an important difference.

question is ''for the same sort of reason as given a second ago: though moral responsibility and metaphysical freedom are not equivalent notions they are conceptual cousins''. If one uses Frankfurtian considerations about PAP to argue for R-Compatibilism and then combines that result with reflections on the intimate connection between metaphysical freedom and moral responsibility, one might naturally be led to an argument for M-Compatibilism. Let's investigate what this argument involves.

The conclusion of the argument needs to be that M-Compatibilism is true. Furthermore, as we just observed, the argument begins by using Frankfurtian considerations to support R-Compatibilism. So the first premise of the argument is surely that moral responsibility is consistent with causal determinism and the premise is supported by reflection on Frankfurt cases. If one wants to tie this premise to the M-Compatibilist conclusion, there seems to be an obvious way to do this that is worth exploring. What *is* needed, and *all* that is needed, to argue from R-Compatibilism to M-Compatibilism is the thesis that moral responsibility *requires* metaphysical freedom: equivalently, that metaphysical freedom is a necessary condition for moral responsibility. Let's now display the full argument and let's call it the ''Freedom/Responsibility Argument''. Here it is:

The Freedom/Responsibility Argument

P1. Moral responsibility is consistent with causal determinism. = R-Compatibilism

P2. Moral responsibility requires metaphysical freedom.

C1. So, metaphysical freedom is compatible with causal determinism. = M-Compatibilism

Note that the conjunction of the argument's first premise with its conclusion is equivalent to Full Compatibilism. I suspect that many philosophers who are Full Compatibilists and arrive at their position from a starting point of reflections on Frankfurt's work on PAP would, if pressed, formulate this very argument in defense of their Full Compatibilist position.[5] The argument is also of independent interest even if no contemporary compatibilist would endorse it. The argument is at

[5] There are other ways to arrive at the Full Compatibilist position. One might do so, for example, via independent reflections on Frankfurt on PAP (supporting R-Compatibilism) and some independent argument for M-Compatibilism. For some critical discussion of independent arguments for M-Compatibilism see van Inwagen (1983) and Warfield (2003).

least initially plausible and suggests that the two compatibilism debates are tightly enough connected to rule out some contemporary positions (e.g. semi-compatibilism). It is time to examine this argument more closely.

II. Evaluation of the Freedom/Responsibility Argument

The Freedom/Responsibility argument is formally valid: if A is consistent with C and A implies B then B is consistent with C. Because the argument is driven primarily by the Frankfurtian considerations in favor of its first premise, I will devote the bulk of my attention to this premise. Before beginning that discussion, let me say a tiny bit more about the argument's second premise. This premise was chosen because it is the natural (and I think needed) way to get from R-Compatibilism to M-Compatibilism and because this is one natural suggestion about the nature of the alleged ''intimate connection'' between moral responsibility and metaphysical freedom. These are good reasons to *consider whether* the premise is true, but why should we *believe that* the premise is true?

Let me make two points for now. We will return to this premise later. First, one good thing for fans of the Freedom/Responsibility argument is that at least many M-Incompatibilists (the primary targets of the argument) will concede the premise immediately. Most M-Incompatibilists will agree with those defending the argument that the ''real action'' with this argument is in the evaluation of the first premise. Secondly, we do not have to strain ourselves to find at least some intuitive support for the premise. Surely an at least plausible defense against a charge of moral responsibility for some action or outcome would be that ''I did not do it freely''. This, at a minimum, suggests that the claim that moral responsibility requires metaphysical freedom has considerable intuitive appeal.

As we turn to the key first premise of the argument we reach familiar and heavily traveled roads in contemporary philosophy of action. We reach the border of the vast ''Frankfurt debate''. Elsewhere I have contributed to parts of this large discussion.[6] On this occasion I will restrict my attention to those parts of the Frankfurt debate that bear

[6] See Warfield (1996, 2003, and forthcoming). One wanting a good overview of this large set of related issues can do no better than begin with Fischer (1999).

on the evaluation of the Freedom/Responsibility argument. I want to discuss some familiar and some less familiar issues from the broader Frankfurt debate through the lens of the Freedom/Responsibility argument.

What in Frankfurt's work on PAP is supposed to support the claim that moral responsibility is consistent with causal determinism? Clearly it is the ''Frankfurt stories'' that are supposed to do this. Frankfurt stories (I will tell one in a moment) are stories that at least strongly suggest that moral responsibility is consistent with a lack of alternative possibilities.[7] Many think that causal determinism threatens moral responsibility *precisely by* precluding alternative possibilities. So if Frankfurt cases show that the removal of alternative possibilities does not thereby remove moral responsibility, then Frankfurt stories provide at least some reason to think that R-Compatibilism is true.

Let's look at one Frankfurt story:

> Larry rudely pushes Freddy. Larry pushes Freddy because this is what Larry wants to do. Larry is in no way pressured into pushing Freddy. It seems quite plausible to say in this case that Larry is responsible for pushing Freddy. This is so even if we add to the story that Geoffrey was monitoring Larry's activities and stood ready and able to force Larry to push Freddy if Larry showed any inclination not to do so on his own.

Because Larry pushed Freddy ''on his own''—that is, without Geoffrey's intervention—and did so precisely because he wanted to do so, it is at least tempting to say that Larry is morally responsible for pushing Freddy. Larry, it is natural to say, would have pushed Freddy even if Geoffrey had been absent from the scene. It is also tempting to claim that because of the presence, plan, and ability of Geoffrey as stipulated in the story, there were no alternative possibilities to Larry's pushing Freddy: it was going to happen no matter what. Perhaps then, this is a case of moral responsibility without alternative possibilities and therefore a case which provides, as discussed in the preceding paragraph, some reason to think that causal determinism is consistent with moral responsibility.

If this reasoning in support of R-Compatibilism (equivalently, premise one of the Responsibility/Freedom argument) is allowed to stand along

[7] Some think the stories somehow literally show that this is correct but this is surely too strong a claim (if it's a coherent claim at all).

with the earlier support for the second premise of the argument then M-Incompatibilists of both stripes (both Full Incompatibilists and Semi-Compatibilists) are in trouble. This argument drives straight through from R-Compatibilism to M-Compatibilism which means that if the argument is sound we are stuck with Full Compatibilism. The argument can and should be resisted.

I will now present and discuss three big worries about the Freedom/Responsibility argument. Two of these worries are about the support just discussed for the argument's first premise. The third objection concerns the interplay of the argument's two premises. Some, but not all, of these points will be at least somewhat familiar from the broader Frankfurt literature. I will discuss the objections at length.

First, it is at least not obvious that the Frankfurt story we are considering really is a case of moral responsibility without alternative possibilities. I grant, at least for purposes of discussion, the responsibility attribution but challenge the claim that the case does not involve alternative possibilities. After all there are, intuitively, two ways the pushing of Freddy by Larry could have unfolded. In one possible scenario (the actual one) Larry pushes Freddy on his own. In what appears to be an open alternative possible scenario, Larry pushes Freddy only after first showing signs of refraining from doing so and only after being forced by Geoffrey to do so. There seem to be alternative possibilities present in the story both in what the story tells us about Larry and what it tells us about Geoffrey.[8]

Of course, fans of Frankfurt's attack on PAP are likely to want to defend Frankfurt in the following way. Though there may well be ''alternative possibilities'' in some sense in the story we are examining, there are no alternative possibilities to *Larry pushes Freddy*. And though there are no alternative possibilities to this in the story, Larry is, I have seemingly conceded, morally responsible for *Larry pushes Freddy*. If these claims are correct then we do have a case of moral responsibility for an act without alternative possibilities to that act.

This reply distinguishes two readings of PAP. On the first reading, PAP requires that moral responsibility for P requires alternative possibilities

[8] The apparent alternative possibilities involving Geoffrey are probably inessential to the story: the role of Geoffrey could be played by a device deterministically responding to differential inputs from Larry. Though this is a topic of some controversy in the broader Frankfurt literature (see Fischer 1999 for an introduction) it does not appear that the apparent alternative possibilities involving Larry are inessential to the story.

(full stop). On the second reading, PAP requires that moral responsibility for P requires alternative possibilities *to P*. This second reading implies, but is not implied by, the first reading. The response suggested above to my first worry about the support for the Freedom/Responsibility argument's first premise is equivalent to insisting that PAP be interpreted according to this second reading and urging that the Frankfurt story is a clear counter-example to PAP so interpreted.

One wanting to resist this pro-Frankfurt reply might focus (as many have in the broader Frankfurt literature) on promising and metaphysically interesting issues about the individuation of, in this case, acts of pushing. One might suggest, for example, that the pushing of Freddy that would have occurred had Larry not pushed Freddy on his own is an act distinct from the pushing that actually occurred. One might thereby reject the claim that there was no alternative possibility present in the story to the specific act for which Larry seems to be morally responsible, namely, the particular pushing of Freddy that occurred.[9] In the context of the evaluation of the Freedom/Responsibility argument, however, an even better strategy is available. The clarification offered in this defensive reply to my first worry about the Freedom/Responsibility argument leaves the argument vulnerable to a second criticism.

The second criticism is this. Understanding PAP in the way suggested in the response to the first criticism seems to imply that the Frankfurt story does not provide any evidence at all for R-Compatibilism which is, you will recall, the content of the premise that the story is supposed to support. After all, the reply concedes that there are alternative possibilities in the story (just not alternative possibilities to Larry's pushing Freddy). What was supposed to be a Frankfurt ''no alternative possibilities story'' not only does not entail determinism but also, when understood in this way, is not even *consistent with* determinism. We already knew that there being no alternative possibilities in the story does not imply determinism, but we were supposed to take the story as evidence for R-Compatibilism because of reflections on the idea that

[9] In broad terms, this is the ''fine-grained'' response to Frankfurt preferred by, among others, Peter van Inwagen. See, for example, van Inwagen 1983. Defenders of Frankfurt are, of course, aware of this strategy and it has received much attention in the relevant literature. I do not discuss this literature here because in the context of the evaluation of the Freedom/Responsibility argument I think that R-Incompatibilists can pursue an even more promising strategy. Again, see Fischer (1999) for one strong introduction to large parts of this debate.

causal determinism threatens moral responsibility *precisely by* taking away alternative possibilities. Surely, however, the threat causal determinism brings to moral responsibility comes (if it comes at all) in part because causal determinism takes away *all* alternative possibilities.[10] The Frankfurt story as we are understanding it now is not even consistent with causal determinism and so the story seems to have lost all interesting contact with the issue of R-Compatibilism. The Frankfurt story was supposed to be the evidence for R-Compatibilism but on this reading it cannot play that role.[11]

Some R-Compatibilists would dispute this claim. Some defenders of R-Compatibilism think that Frankfurt stories support R-Compatibilism even granting that the stories contain alternative possibilities. Most notably, this is the position of John Martin Fischer (see Fischer 1994 and 1999 and Fischer and Ravizza 1998). Fischer claims that the alternative possibilities identified in Frankfurt stories are ''insufficiently robust'' to ''ground'' the moral responsibility we find in the stories. He concludes that the stories support R-Compatibilism despite the presence of alternative possibilities in them. ''Grounding'', however, is surely a matter of *sufficiency*, not *necessity*. No one is claiming that alternative possibilities are sufficient for moral responsibility: R-Incompatibilists claim only that alternative possibilities are necessary for moral responsibility. R-Compatibilists can therefore agree with Fischer that the alternative possibilities do not ''ground'' the responsibility that appears to be present in Frankfurt cases. It is not easy to see how this discussion of ''grounding'' is supposed to help R-Compatibilism.[12]

R-Compatibilists who concede that the Frankfurt story contains alternative possibilities seem to encounter difficulties. Can R-Compatibilists

[10] A small minority of philosophers (holding various positions on the two compatibilism debates) do not think that determinism threatens moral responsibility by taking away alternative possibilities. These philosophers think that even *determinism* is consistent with the presence of genuine alternative possibilities (some use this claim in attempts to support R- and/or M-Compatibilism but not all do). I set aside this minority position for purposes of this chapter.

[11] This reply grants to the Frankfurtian the second reading of PAP noted above and the claim that the Frankfurt story provides a counter-example to it. This reply, however, emphasizes that a counter-example to this second reading of PAP provides no support for R-Compatibilism.

[12] Fischer is likely to claim further that if we concede that the alternative possibilities do not ground moral responsibility in these stories, then the alternative possibilities also cannot *make the difference* between cases with moral responsibility and cases without moral responsibility. This latter claim does not follow: small differences often make big differences in philosophical analyses. This debate about the ''robustness'' of alternative

stipulate that causal determinism obtains in the Frankfurt story and thereby eliminate the presence of alternative possibilities in the story? If so, the story and its evaluation are clearly relevant to the discussion of both R-Compatibilism and M-Compatibilism. This attempt to stipulate the relevance of the story, however, runs into severe difficulties. If one adds to our Frankfurt story that Larry's pushing of Freddy occurs in a causally deterministic situation with literally no alternative possibilities, one must surely say more about the two places in the Frankfurt story that seem to contain genuine alternative possibilities.

One making this stipulation would need to explain in what sense, if any, it was open to Larry to push Freddy *not* on his own (as in the actual story) but only after being forced to do so by the intervention of the inactive counterfactual intervener Geoffrey. And one would need to say more about the sense in which inactive counterfactual intervener Geoffrey stood by ready and *able* to intervene. Both parts of the story seem important for defending the key claims that Larry is responsible though he could not have done otherwise (though perhaps only Larry's apparent alternative possibilities are essential to it). Supposing Larry and Geoffrey to be able to do other than they actually do in a causally deterministic story is to suppose that M-Compatibilism is true. As part of the Freedom/Responsibility argument, however, the story is supposed to be supporting a premise of an overall *argument for M-Compatibilism* and so this supposition is presumably not permissible in this setting. I therefore fail to see how the Frankfurt story is supposed to support R-Compatibilism and fail to see how the story is supposed to support the first premise of the Freedom/Responsibility argument.

I conclude that those defending R-Compatibilism by appealing to Frankfurt are in some trouble. They need some response to the first worry I raised about the defense of the Freedom/Responsibility argument's first premise. In the spirit of philosophical cooperation I offered them a response to this worry. The response I generously provided may well be a strong defense of *Frankfurt's rejection of one reading of PAP*. As I have just pointed out, however, the response is of no use to one wanting to appropriate Frankfurt's attack against PAP *in a defense of the Freedom/Responsibility argument*. The Full Compatibilist therefore

possibilities in Frankfurt stories is, of course, an ongoing debate. I do not claim to have ended the debate here. See the Fischer references already cited and Warfield (2003 and forthcoming) for further discussion.

needs to look for a different response to the first worry I raised if she is to avoid the second worry. I have no suggestion to make on her behalf.

I turn now to my third and final worry about the Freedom/Responsibility argument. Though I have argued that this is unlikely, assume for purposes of discussion that we somehow come to see that Frankfurt cases strongly support R-Compatibilism. Even stronger, assume for purposes of discussion that R-Compatibilism is true. Are we then pulled inevitably to Full Compatibilism? I think that the answer to this question is ''no''. Assuming the truth of R-Compatibilism helps us focus on worries we should have about the second premise of the Freedom/Responsibility argument.

Let's take another look at this second premise. If we really did accept that the Frankfurt story involves moral responsibility without alternative possibilities in some way that helps establish R-Compatibilism, M-Incompatibilists should challenge the second premise of the argument using the very Frankfurt story introduced by the R-Compatibilist. That is, M-Incompatibilists should use Frankfurt stories to defend Semi-Compatibilism.

Recall the description of Larry and his behavior. Larry pushed Freddy, did so because he wanted to, and was not pressured or forced to do so by Geoffrey or anyone else. The M-Incompatibilist will surely note that none of this is *sufficient* for Larry's being metaphysically free in the story. Perhaps this description *is* sufficient for Larry's being metaphysically free in the story *if* M-Compatibilism is true. But again, the premise under discussion is part of an argument for M-Compatibilism and so it would be dialectically inadvisable to introduce M-Compatibilism as an additional premise supporting premise two of the original argument.

The M-Incompatibilist observation here is simply this: the Frankfurt story is at least as plausible when considered as a candidate for being a story in which Larry is morally responsible *but not metaphysically free*, as it is when viewed as a candidate for being a story in which Larry is morally responsible but lacks alternative possibilities. The more we incline towards accepting the first premise of the Freedom/Responsibility argument, the more we are pushed towards rejecting the argument's second premise (or so the M-Incompatibilist should say). At one limit of this relation, if we concede for purposes of discussion that the story provides perfect support for the first premise, we should insist that the

story also provides perfect support for the rejection of the second premise.

Can a defender of the Freedom/Responsibility argument point to the existence of alternative possibilities in the story as evidence of Larry's metaphysical freedom? This would support the claim that the Frankfurt story is a story involving both moral responsibility and metaphysical freedom, consistent with the second premise of the Freedom/Responsibility argument. The Freedom/Responsibility argument's defender would, however, be ill-advised to make this move. The M-Incompatibilist would be too quick to embrace the point. The point concedes the earlier criticisms of the first premise of the argument, conceding that there are alternative possibilities present in the story that is supposed to be a case involving moral responsibility without alternative possibilities.

III. CONCLUSION

Lots of philosophers are compatibilists about causal determinism and both metaphysical freedom and moral responsibility. One source of this Full Compatiblism is, oddly enough, Harry Frankfurt's work on PAP. This is odd because, as noted earlier, Frankfurt's work on PAP directly concerns neither metaphysical freedom nor causal determinism and therefore makes explicit contact with only one of the three concepts in play in contemporary discussions of Full Compatibilism. It is, however, easy to see how to transition from Frankfurt's work on PAP to R-Compatibilism and then on to M-Compatibilism. The Freedom/Responsibility argument brings out this natural progression. I hope to have convinced you that there are some weaknesses and serious confusions in this argument.

Though Frankfurt's attack on PAP can be resisted, it can also be reasonably endorsed. Philosophers can reasonably go either way on this issue. Those who go further and appropriate Frankfurt for purposes of defending R-Compatibilism are, I believe, making a mistake, but their position is also defensible and it is not crazy to defend it in this way. Metaphysicians who press on from these positions in an attempt to reach M-Compatibilism (and therefore Full Compatibilism) have, it

seems to me, stretched the resources of the Frankfurt material beyond the breaking point.[13]

REFERENCES

Fischer, J. M. (1994) *The Metaphysics of Free Will* (Blackwell).
——(1999) ''Recent Work on Moral Responsibility'', *Ethics* 110: 93–139.
——and M. Ravizza, (1998) *Responsibility and Control* (Cambridge University Press).
Frankfurt, H. (1969) ''Alternative Possibilities and Moral Responsibility'', *Journal of Philosophy* 66.
van Inwagen, P. (1983) *An Essay on Free Will* (Oxford University Press).
Warfield, T. A. (1996) ''Determinism and Moral Responsibility are Incompatible'', *Philosophical Topics* 24.
——(2000) ''Causal Determinism and Human Freedom are Incompatible: A New Argument for Incompatibilism'', *Philosophical Perspectives* 14.
——(2003) ''Compatibilism and Incompatibilism: Some Arguments'', in M. Loux and D. Zimmerman (eds.) *The Oxford Handbook of Metaphysics* (Oxford University Press).
——(forthcoming) *Metaphysical Freedom*.

[13] I thank Thad Botham, Tom Crisp, EJ Coffman, John Martin Fischer, Christian Miller, Mike Rea, and Ted Sider for helpful discussion. An earlier version of this paper was presented at the University of Colorado where the general discussion was enjoyable and where Michael Tooley, Michael Huemer, Christopher Shields, and George Bealer provided insightful commentary. Dean Zimmerman and an anonymous referee for *Oxford Studies in Metaphysics* provided extremely helpful organizational and substantive feedback on the penultimate draft. As always when working on this topic, my deepest intellectual debts are to Peter van Inwagen and Christopher Hill.

This page intentionally left blank

INDEX OF NAMES